

Debreceni Egyetem  
Informatikai Kar

## Iterációs Lemmák a Formális Nyelvek Elméletében

Témavezető:  
Dr. Dömösi Pál

Készítette:  
Hegedüs László  
Programtervező Informatikus (Bsc)

Debrecen  
2009

# Tartalomjegyzék

|                                                                       |           |
|-----------------------------------------------------------------------|-----------|
| <b>1. Bevezetés</b>                                                   | <b>1</b>  |
| <b>2. Matematikai alapfogalmak, jelölések</b>                         | <b>3</b>  |
| 2.1. Halmazok, halmazműveletek . . . . .                              | 3         |
| 2.2. Relációk, leképezések . . . . .                                  | 4         |
| 2.2.1. Relációk . . . . .                                             | 4         |
| 2.2.2. Leképezések . . . . .                                          | 5         |
| 2.3. Gráfok, fák . . . . .                                            | 5         |
| 2.3.1. Gráfok . . . . .                                               | 5         |
| 2.3.2. Fák . . . . .                                                  | 6         |
| 2.4. Algebrai alapfogalmak . . . . .                                  | 6         |
| 2.4.1. Algebrai struktúrák . . . . .                                  | 6         |
| <b>3. Ábécé, szavak, nyelvek</b>                                      | <b>8</b>  |
| 3.1. Nyelvek, műveletek nyelvekkel . . . . .                          | 9         |
| <b>4. Átírási rendszerek, generatív nyelvtanok</b>                    | <b>10</b> |
| 4.1. Átírási rendszerek . . . . .                                     | 10        |
| 4.2. Generatív nyelvtanok . . . . .                                   | 10        |
| <b>5. Reguláris nyelvek</b>                                           | <b>13</b> |
| 5.1. Reguláris halmazok és reguláris kifejezések . . . . .            | 13        |
| 5.2. Véges automaták . . . . .                                        | 13        |
| 5.3. Véges automaták megadása . . . . .                               | 18        |
| 5.4. Reguláris nyelvek tulajdonságai . . . . .                        | 19        |
| <b>6. Környezetfüggetlen nyelvek</b>                                  | <b>21</b> |
| 6.1. Veremautomaták . . . . .                                         | 21        |
| 6.2. Környezetfüggetlen nyelvek tulajdonságai, átalakításai . . . . . | 22        |
| 6.3. Veremautomaták megadása . . . . .                                | 25        |
| 6.4. Derivációs fák, levezetések . . . . .                            | 25        |
| 6.5. Lineáris nyelvek . . . . .                                       | 26        |
| <b>7. Iterációs lemmák</b>                                            | <b>27</b> |
| 7.1. Iterációs lemmák reguláris nyelvekre . . . . .                   | 27        |
| 7.2. Iterációs lemmák környezetfüggetlen nyelvekre . . . . .          | 29        |

|                               |           |
|-------------------------------|-----------|
| <b>8. Összefoglalás</b>       | <b>41</b> |
| <b>9. Köszönetnyilvánítás</b> | <b>43</b> |

# 1. Bevezetés

A formális nyelvek és automata elmélete egy dinamikusan fejlődő tudományág. Az algoritmusok, fordítóprogramok, szövegfeldolgozás, új számítási paradigmák és még sok más terület elméleti alapjait képezi. Az automataelmélet az 1920-1930 -as években alakult ki, ami elvezetett nem sokkal később, az 1950 -es években megjelenő formális nyelvek elméletéhez. Ha pontos évszámot akarnánk mondani, akkor 1956 -ban, Noam Chomsky publikációinak hatására történt. Ő 4 nyelvosztályba csoportosította a formális nyelveket. Ezek után az egyik legjellemzőbb probléma: eldönteni egy nyelvről, hogy egy adott nyelvosztályba tartozik -e, vagy sem. A megoldásra több módszer is létezik: a nyelvosztályok zártági tulajdonságait felhasználva vizsgáljuk az adott nyelvet, vagy kicserélési lemmákat alkalmazunk, de a legelterjedtebb és általában a legegyszerűbb módszer iterációs lemmák alkalmazása. Dolgozatomban az utóbbira térek ki részletesen, ismertetem a témával kapcsolatos főbb fogalmakat, az elterjedtebb lemmákat, hiányosságaikat és különböző javításaikat. Különös tekintettel a környezetfüggetlen és reguláris nyelvekre, mivel ezen nyelvosztályok bírnak a legnagyobb gyakorlati jelentőséggel. A nyelvek jellemzésénél kitérek az őket felismerő automatákra, fontosabb tulajdonságaikra, valamint a nyelvek zártági tulajdonságaira.

A 2. fejezetben ismertetem a dolgozatban használt matematikai alapfogalmakat, jelöléseket: halmazok és rajtuk végezhető műveletek, relációk, leképezések, gráfok és fák. A 3. fejezetben a formális nyelvek alapfogalmai szerepelnek: ábécé, szavak, szavakon végezhető műveletek, nyelvek, műveletek nyelvekkel. A generatív nyelvtan fogalmával a 4. fejezetben találkozunk, az 5. és a 6. fejezetekben pedig rendre a reguláris- és környezetfüggetlen nyelvek leírása található. Végül a 7. fejezetben foglalkozunk az egyes elterjedtebb iterációs lemmák vizsgálatával.

Az első iterációs lemma, mely Bar-Hillel lemma néven vált ismertté, Yehoshua Bar-Hillel, M. Perles, és E. Shamir nevéhez fűződik, akik 1961 -ben publikálták cikküket [3]. Ezen lemma segítségével érthetően és elegánsan be lehetett bizonyítani, hogy egyes nyelvek nem környezetfüggetlenek. További iterációs lemmák jelentek meg különböző nyelvosztályokhoz, azonban kijelenthetjük szinte az összesről, hogy adott nyelvosztályba való tartozásnak csak az elégséges feltételeit vizsgálják. Tehát, ha egy nyelv teljesíti egy adott lemma, például a Bar-Hillel lemma feltételeit, még nem jelenti azt, hogy a nyelv környezetfüggetlen. Ekkor egy megoldás az, hogy szűkítjük a lemma feltételeit. Így tett W. Ogden 1968 -ban megjelent cikkében [9]. Ogden lemmájának segítségével továbbá azt is meg lehet mutatni, hogy egy adott nyelvtan természeténél fogva kétértelmű, vagyis bármely, a nyelvtan által generált szónak legalább két legbaloldalibb levezetése is létezik. A dolgozatban erre nem térek ki. Wise 1976 -os lemmája a környezetfüggetlen nyelvek osztályába tartozásnak szükséges és elégséges feltételét is vizsgálja [13]. Ezzel a lemmával bonyolultsága miatt nem foglalkozunk a dolgozatban. Ogden lemmájának szűkített

feltételeit fogalmazta meg C. Bader és A. Mora 1982 -es publikációjukban [2]. Dömösi Pál és Manfred Kudlek több nyelvosztályra vonatkozó erős iterációs lemmákat ismertetett 1999 -es cikkükben [5]. Néhány évtized alatt jelentős számú lemma jelent meg, mindegyik valamivel több megszorítást tartalmazott, mint az őket megelőzők. Ugyanakkor a legtöbb esetben akár a legegyszerűbb lemmák alkalmazásával is sikeresen be lehet bizonyítani a feltevéseinket.

## 2. Matematikai alapfogalmak, jelölések

Ebben a fejezetben ismertetem a dolgozat további részében használatos alapvető matematikai fogalmakat, jelöléseket.

### 2.1. Halmazok, halmazműveletek

Halmaznak nevezzük valamilyen objektumok gyűjteményét, ezeket az objektumokat pedig a halmaz elemeinek hívjuk. Ha egy adott  $x$  elemet egy  $A$  halmaz tartalmaz, az  $x \in A$  jelölést alkalmazzuk. Fontos megjegyezni, hogy egy halmaz elemei között semmilyen sorrendet nem tételezünk fel. Két halmaz akkor és csak akkor egyenlő, ha valahányszor az egyik tartalmaz egy elemet, akkor az az elem a másik halmaznak is eleme.  $A = B \Leftrightarrow$  minden  $x \in A$  -ra  $x \in B$  és minden  $y \in B$  -re  $y \in A$  is teljesül.

Halmazokat megadhatunk az elemek felsorolásával, például:

$$\{1, 2, a, b\},$$

vagy megadjuk, hogy az elemekre milyen feltételnek kell teljesülnie:

$$\{x | x \in \mathbb{N}, x \text{ -páros}\}.$$

Egy  $A$  halmaz számosságán az  $A$  -ban lévő összes elem darabszámát értjük. Jelölése:  $|A|$ . Így a halmazok két csoportba sorolhatók: véges és végtelen számosságú halmazok. Végtelen számosságú halmazok például a természetes számok halmaza,  $\mathbb{N}$ , vagy a valós számok halmaza,  $\mathbb{R}$ . Véges számosságú halmaz például a magyar ábécé magánhangzóit tartalmazó halmaz. Egy speciális halmaz az üres halmaz, melynek egyetlen eleme sincs, az üres halmaz jele:  $\emptyset$ . A következő definícióban vannak megfogalmazva a halmazokon végezhető alapműveletek.

**2.1. Definíció.** Legyen  $A$  és  $B$  halmaz. Ekkor:

- $A$  és  $B$  unióján azokat az elemeket értjük, melyek  $A$  -nak, vagy  $B$  -nek elemei. Jelöléssel:

$$A \cup B = \{x | x \in A, \text{ vagy } x \in B\}$$

- $A$  és  $B$  metszete azokat az elemeket tartalmazza, melyek  $A$  -nak és  $B$  -nek is elemei. Jelöléssel:

$$A \cap B = \{x | x \in A \text{ és } x \in B\}$$

- $A$  és  $B$  halmazok különbségét azok az elemek alkotják, melyek  $A$  -nak elemei, de  $B$  -nek nem. Jelöléssel:

$$A \setminus B = A - B = \{x | x \in A \text{ és } x \notin B\}$$

- $A$  és  $B$  halmazok Descartes-szorzatát olyan  $(x, y)$  párok alkotják, melyekben  $x \in A$  és  $y \in B$ . Jelöléssel:

$$A \times B = \{(x, y) | x \in A, y \in B\}$$

Legyenek  $A$  és  $B$  halmazok.  $B$  akkor és csak akkor részhalmaza  $A$ -nak, ha  $B$  összes eleme  $A$ -nak is eleme. Jelöléssel:

$$B \subseteq A \Leftrightarrow x \in A \quad \text{minden } x \in B \text{-re.}$$

Továbbá  $B$  akkor és csak akkor valódi részhalmaza  $A$ -nak, ha  $A$ -nak  $B$  elemein kívül vannak más elemei, tehát:

$$B \subset A \Leftrightarrow B \subseteq A \text{ és } |B| < |A|.$$

Egy  $A$  halmaz összes részhalmazának a halmazát  $2^A$ -val jelöljük. Az üres halmaz minden halmaznak részhalmaza.  $2^A$  számossága  $2^{|A|}$ .

Legyen  $A$  egy halmaz és  $B \subset A$ .  $B$  komplementerén azokat az  $A$ -beli elemeket értjük, melyek  $B$ -nek nem elemei. Jelöléssel:

$$\overline{B} = \{x | x \in A \text{ és } x \notin B\}.$$

Ha tetszőleges  $A$  és  $B$  halmazokra teljesül, hogy  $A \cap B = \emptyset$ , akkor azt mondjuk, hogy  $A$  és  $B$  diszjunkt halmazok.

## 2.2. Relációk, leképezések

### 2.2.1. Relációk

Adott  $X_1, X_2, \dots, X_n$  halmazok felett értelmezett reláció a következő rendezett  $n$ -es:

$$\varrho \subseteq (X_1 \times X_2 \times \dots \times X_n).$$

Pontosabban megfogalmazva ez egy  $n$  változós reláció. Fontos, hogy  $X_1, X_2, \dots, X_n$  halmazok sorrendje rögzített. Tehát például  $n = 3$  esetén  $(x_1, x_2, x_3) \neq (x_2, x_1, x_3)$ , ahol  $x_i \in X_i$  minden  $i = 1, 2, 3$ -ra. Ha  $n = 1, 2, 3, \dots$ , akkor rendre unáris, bináris, ternáris, ... stb. relációról beszélünk. A dolgozatban részletesebben csak a bináris relációkkal foglalkozunk. Egy bináris relációt tehát  $\varrho = (X_1, X_2)$  párossal jelölünk. Legyen  $x \in X_1$  és  $y \in X_2$ , ha  $x$  relációban áll  $y$ -nal, akkor  $(x, y) \in \varrho$  teljesül, melyet jelölhetünk  $x \varrho y$ -al is. Gyakran előfordul, hogy  $X_1 = X_2$ , ekkor, ha egyértelmű, hogy bináris relációról van szó, elég a halmazt csak egyszer feltüntetni.

**2.2. Definíció** (Bináris relációk tulajdonságai). Egy  $\varrho(A)$  bináris reláció:

- reflexív, ha  $a\varrho a$  fennáll minden  $a \in A$ -ra, egyébként nem reflexív. Ha nem létezik olyan  $a \in A$ , amelyre  $a\varrho a$  fennáll, akkor  $\varrho$  irreflexív.
- tranzitív, ha  $a\varrho b$  és  $b\varrho c$  fennállásából következik, hogy  $a\varrho c$  is teljesül.
- szimmetrikus, ha  $a\varrho b$  esetén  $b\varrho a$  is fennáll. Ha  $a\varrho b$  esetén  $b\varrho a$  biztos, hogy nem áll fenn, akkor  $\varrho$  antiszimmetrikus.

Ha egy  $\varrho(A)$  bináris reláció reflexív, szimmetrikus és tranzitív, akkor  $\varrho$   $A$ -n értelmezett ekvivalencia reláció. A  $\varrho$  reláció  $A$ -t diszjunkt ekvivalencia osztályokra (halmazokra) bontja, melyeket jelöljünk  $A_1, A_2, \dots, A_n$ -el, ahol  $n = 1, 2, \dots$ . Könnyen belátható az előbbi kijelentés, mivel  $\varrho$  tranzitív és szimmetrikus tulajdonsága miatt könnyen külön halmazokba szervezhetjük az elemeket. Kiindulásként  $A$  valamely eleme legyen  $A_1$  egyetlen eleme, majd minden  $b \in A$ -ra, ha  $a\varrho b$  teljesül, áthelyezzük  $b$ -t  $A_1$ -be. Ha végeztünk és  $A$  nem üres, akkor  $A$  egyik eleme most legyen  $A_2$  egyetlen eleme és helyezzük át  $A$ -ból a vele relációban álló elemeket  $A_2$ -be, majd jön az  $A_3$  halmaz és így tovább. Így, ha  $A$  véges, akkor véges sok lépésben az elemei elfogynak. Ha  $A$  végtelen, akkor végtelen sok ekvivalencia osztályt kaphatunk.

### 2.2.2. Leképezések

Egy  $\varphi \subseteq (X, Y)$  bináris relációt leképezésnek hívunk, ha valamely  $x \in X$ -re és  $y \in Y$ -ra  $(x, y) \in \varphi$ , akkor bármely  $y' \in Y$ -ra,  $y' \neq y$ -ból következik, hogy  $(x, y') \notin \varphi$ . Tehát  $\varphi$  az  $X$  halmaz elemeihez egyértelműen rendel párt  $Y$  elemei közül. A relációkat szokás a következőképpen is jelölni:

$$\varphi : X \rightarrow Y.$$

Az  $X$  halmazt a reláció értelmezési tartományának nevezzük,  $Y$  pedig a reláció értékkészlete. Fontos megjegyezni, hogy egy halmaznak is lehetnek halmazok az elemei, ezért a  $\varphi : X \rightarrow 2^Y$  leképezés is helyes, továbbá érték  $n$ -eseket tartalmazó halmazokból, vagy halmazokba történő leképezés is lehetséges. Ennek tisztázására az automatáknál definiált átmenetfüggvény alakja miatt volt szükség.

## 2.3. Gráfok, fák

### 2.3.1. Gráfok

Egy adott gráfot egy  $G = (C, E)$  rendezett páros jellemez, ahol mind  $C$ , mind  $E$  véges halmazok, rendre a csomópontok halmaza és csomópont párokból álló élek halmaza. Két adott  $c$  és  $v$  csomópontot él köt össze, ha  $(c, v) \in E$ . Egy adott csomópontból egy másikba vezető él

sorozatát sétának hívjuk. Egy séta formálisan:

$$c_1, c_2, \dots, c_n \in C \text{ séta, ha } (c_i, c_{i+1}) \in E \text{ minden } i = 1, 2, \dots, n-1 \text{-re.}$$

A séta hossza a csomópontok közötti élek számával egyezik meg, tehát ebben az esetben  $n - 1$ . Útnak nevezzük azt a sétát, amelyben a csomópontok páronként különbözőek.

Az előbb definiált típusú gráfoknál, ha  $c$ -ből vezet él  $v$ -be, akkor ez visszafelé is igaz. Ez nem áll fenn az irányított gráfoknál. Egy  $G = (C, E)$  páros irányított gráf, ahol  $C$  a csomópontok véges halmaza,  $E$  pedig  $(c, v)$  alakú párok halmaza, ahol  $c, v \in C$  és  $(c, v)$  egy irányított élt jelöl, mely  $c$ -ből  $v$ -be mutat. Egy  $c$ -ből  $v$ -be mutató irányított élt szokás a következő módon is jelölni:

$$c \rightarrow v.$$

Azt mondjuk, hogy  $c$  megelőzi  $v$ -t, vagy  $c$   $v$  szülője, ha  $c \rightarrow v \in E$ . Továbbá azt is mondhatjuk, hogy  $v$   $c$  rákövetkezője, vagy gyermeke. Az előbbi jelölést felhasználva, az irányított séta fogalmát definiálhatjuk. Legyenek  $c_1, c_2, \dots, c_n \in C$  csomópontok,  $c_1$ -ből akkor vezet irányított séta  $c_n$ -be, ha  $c_i \rightarrow c_{i+1}$  teljesül minden  $i = 1, 2, \dots, n-1$ -re. Egy irányított séta akkor irányított út, ha a csomópontjai páronként különbözőek. Azt az irányított sétát, melynek kezdő és vég csúcspontja megegyezik, körnek nevezzük.

### 2.3.2. Fák

Az irányított gráfok speciális esetei a fák. A fának van egy kitüntetett eleme, a gyökérelem, melybe egyetlen él sem vezet és belőle minden másik csomóponthoz vezet irányított út. Minden csomópontnak (kivéve a gyökérelemtől) pontosan egy szülője van. Ha egy  $c$  csomópontnak  $p$  a szülője (vagyis  $p$ -ből mutat él  $c$ -be), akkor  $c$  a  $p$  gyermeke. Azokat a csomópontokat, melyekbe vezet út egy  $p$  csomópontból, a  $p$  leszármazottainak hívjuk. A fában speciális elemek a levélelemek, melyeknek egyetlen leszármazottjuk sincs, az összes többi elemet belső csomópontoknak nevezzük.

Egy fa gyökéreleme a fa nulladik szintjén áll, a gyökérelem leszármazottai az első szinten és így tovább. A fa magassága pedig a legmagasabb szintjének a sorszáma.

## 2.4. Algebrai alapfogalmak

### 2.4.1. Algebrai struktúrák

Bináris műveletnek nevezünk egy  $\circ : A \times A \rightarrow A$  leképezést. Műveleteknél az  $(a, b) \in \circ$  jelölés helyett a relációknál ismertetet  $a \circ b$  jelölést szokás alkalmazni. Azt mondjuk, hogy az  $A$  halmaz zárt a  $\circ$  műveletre nézve, ha minden  $a, b \in A$ -ra  $a \circ b \in A$  teljesül. A  $\circ$  művelet:

kommutatív, ha  $a \circ b = b \circ a$  minden  $a, b \in A$ -ra. Valamint  $\circ$  asszociatív, ha  $(a \circ b) \circ c = a \circ (b \circ c)$  minden  $a, b, c \in A$ -ra. A  $\circ$  művelet egységeleme egy  $e \in A$  elem, ha  $a \circ e = e \circ a = a$  minden  $a \in A$ -ra. Az  $a \in A$  elem inverze  $b \in A$  akkor és csak akkor, ha  $a \circ b = b \circ a = e$  teljesül, ahol  $e \in A$  az egységelem.

Félcsoportnak nevezünk egy  $(A, \circ)$  rendezett párt, ahol  $A$  egy nem üres halmaz és  $\circ$   $A$  elemein értelmezett asszociatív bináris művelet. Azt a félcsoportot, mely tartalmaz egységelemet, egységelemes félcsoportnak, más néven monoidnak nevezzük. Csoportnak nevezzük azt az  $(A, \circ)$  monoidot, melyben minden  $a \in A$ -nak létezik inverze és ábel-csoportnak, ha igaz az is rá, hogy a  $\circ$  művelet kommutatív.

### 3. Ábécé, szavak, nyelvek

Ebben a fejezetben ismertetem az alapvető matematikai nyelvészeti jelöléseket, fogalmakat, definíciókat.

Az adatokat, objektumokat általában szimbólumok sorozataként ábrázoljuk. Szimbólumok egy tetszőleges véges, nem üres halmazát ábécének nevezzük. Jelölése:  $V$ . Egy  $V$  ábécé feletti szónak nevezzük  $V$ -ben lévő szimbólumok egy  $w$  véges sorozatát. Formálisan:

$$w = w_1 w_2 \dots w_n, \quad \text{ahol} \quad w_1, w_2, \dots, w_n \in V$$

Továbbá  $w$  szó hosszán  $w$ -t alkotó összes szimbólum darabszámát értjük (beleértve a duplikált szimbólumokat is). Jelölése:  $|w|$ . A fentebbi  $w$  szó hossza  $n$ , vagyis  $|w| = n$ . Ezek alapján beszélhetünk  $w$ -ben  $i$ -edik pozícióban lévő szimbólumról, ahol  $i = 1, 2, \dots, |w|$ . Azt a szót, amely egyetlen szimbólumból sem áll, üresszónak nevezzük és  $\lambda$ -val jelöljük. Könnyen látható, hogy  $|\lambda| = 0$ .

Egy  $V$  ábécé feletti összes véges szavak (köztük az üres szó is) halmazát  $V^*$ -al jelöljük ( $V^*$ -ot szokás  $V$  Kleene-lezártjának is nevezni).  $V$  feletti összes nem üres véges szavak hamazának jele:  $V^+$ . ( $V^+ = V^* - \{\lambda\}$ ). Véges szó alatt véges hosszúságú szavakat értünk.

Egy  $x \in V^*$  szóban az  $a \in V$  előfordulásainak számát  $|x|_a$ -val jelöljük. Egy adott  $w \in V^*$ ,  $w = w_1 w_2 \dots w_n$  szó tükörképe:

$$w^R = w_n w_{n-1} \dots w_1.$$

Egy  $V$  ábécé feletti  $i \geq 0$ -nál nem hosszabb szavakat a következő módon jelöljük:  $V^{(i)}$ ,  $x \in V^{(i)}$  akkor és csak akkor, ha  $|x| \leq i$ .

Egy  $V$  ábécé feletti szavakon értelmezett szorzás (konkatenáció) művelete a következő:

$$K : V^* \times V^* \rightarrow V^*$$

$$x = x_1 x_2 \dots x_n, \quad y = y_1 y_2 \dots y_m$$

$$K(x, y) = x \cdot y = x_1 x_2 \dots x_n y_1 y_2 \dots y_m, \quad \text{minden } x, y \in V^* \text{-ra.}$$

Tehát  $u, v$  szavak szorzata a két szó egymás után írása.

A szorzás művelete asszociatív  $V^*$ -ra nézve:  $x \cdot (y \cdot z) = xyz = (x \cdot y) \cdot z$  minden  $x, y, z \in V^*$ -ra, továbbá minden  $x \in V^*$ -ra teljesül:  $x \cdot \lambda = \lambda \cdot x = x$ . Tehát  $V^*$  a szorzással monoidot alkot, ahol  $\lambda$  egységelem.

A konkatenáció műveletét felhasználva minden  $V$  ábécé feletti  $x$  szóra definiálhatjuk  $x$   $i$ -edik hatványát a következőképpen:

$$x \in V^*, \quad x^i = x \cdot x^{i-1}, \quad \text{ahol } i \geq 1 \text{ és } x^0 = \lambda.$$

Továbbá  $x^*$  is értelmezett, ami  $x$  tetszőleges sokszor (akár nullszor is) egymás után írását jelenti. Erre a jelölésre majd a reguláris kifejezéseknél visszatérünk.

Legyenek  $v, w \in V^*$  két  $V$  ábécé feletti szó. Defináljuk a rész szó fogalmát a következőképpen:  $v$  akkor és csak akkor rész szava  $w$  -nek, ha léteznek  $x, y \in V^*$  szavak úgy, hogy  $w = xvy$ .  $v$  valódi rész szava  $w$  -nek, ha  $0 < |v| < |w|$ . A rész szó egyik speciális esete a prefix:  $v$  akkor és csak akkor prefixe  $w$  -nek, ha létezik  $x \in V^*$  szó úgy, hogy  $w = vx$ .  $v$  valódi prefixe  $w$  -nek, ha  $0 < |v| < |w|$ . Másik speciális esete a szuffix:  $v$  akkor és csak akkor szuffixe  $w$  -nek, ha létezik  $x \in V^*$  szó úgy, hogy  $w = xv$ .  $v$  valódi szuffixe  $w$  -nek, ha  $0 < |v| < |w|$ . Például legyen  $w = bokor$ , ekkor  $w$  valódi rész szava lehet  $kor$ , valódi prefixe  $bo$ , valamint valódi szuffixe  $or$ .

### 3.1. Nyelvek, műveletek nyelvekkel

Egy adott  $V$  ábécé feletti szavak egy véges halmazát nyelvnek nevezzük. Jelölése:  $L$ .

$$L \subseteq V^*$$

Egy  $L \subseteq V^*$ ,  $V$  ábécé feletti nyelv komplementere  $\overline{L} = V^* - L$ .

**3.1. Definíció.** Legyen  $L_1, L_2 \subseteq V^*$ .  $L_1, L_2$  nyelveken, mint halmazokon értelmezett az (összeadás) unió-, metszet-, különbségképzés és a (szorzás) konkatenáció művelete:

$$L_1 \cup L_2 = \{x | x \in L_1 \text{ vagy } x \in L_2\}$$

$$L_1 \cap L_2 = \{x | x \in L_1 \text{ és } x \in L_2\}$$

$$L_1 - L_2 = \{x | x \in L_1 \text{ és } x \notin L_2\}$$

$$L_1 L_2 = \{xy | x \in L_1 \text{ és } y \in L_2\}$$

Egy  $L \subseteq V^*$  nyelv  $i$  -edik iteráltja:

$$L^i = \{\lambda\} \cup L \cup LL \cup \dots \cup \overbrace{LL \dots L}^i \quad i \geq 0$$

$L$  tranzitív (Kleene) lezártja

$$L^* = \sum_{i=0}^{\infty} L^i,$$

ahol  $L^0 = \{\lambda\}$ .

## 4. Átírási rendszerek, generatív nyelvtanok

### 4.1. Átírási rendszerek

Egy átírási rendszer (vagy semi-Thue rendszer) olyan  $W = (V, H)$  rendezett pár, ahol  $V$  egy ábécé,  $H$  pedig  $V$  feletti szavak rendezett párpainak halmaza.  $H$  elemeit átírási szabályoknak hívjuk,  $(P, Q)$  alakúak, jelölésük:  $P \rightarrow Q$ .  $P$ -t a szabály bal oldalának,  $Q$ -t a szabály jobb oldalának hívjuk. Legyen  $W = (V, H)$  egy átírási rendszer,  $P, Q \in V^*$ .  $P$ -ből közvetlenül levezethető  $Q$  (vagy  $P$ -ből generálható  $Q$ ), ha létezik  $P', P_1, P'', Q_1 \in V^*$  úgy, hogy  $P = P'P_1P''$  és  $Q = P'Q_1P''$  és  $P_1 \rightarrow Q_1 \in H$ . Ezt  $P \Rightarrow^W Q$ -val jelöljük, de, ha  $W$  egyértelmű, akkor a jelölése elhagyható ( $P \Rightarrow Q$ ).  $P \Rightarrow^* Q$ , azaz  $P$ -ből  $Q$  több lépésben levezethető, ha létezik  $P_0, P_1, P_2 \dots P_k \in V^*$  ( $k \geq 0$ ) szavak véges sorozata úgy, hogy

$$P = P_0, Q = P_k \quad \text{és} \quad P_i \Rightarrow P_{i+1} \quad \text{minden} \quad i = 0, 1 \dots k-1 \text{-re}$$

másképp:

$$P_0 \Rightarrow P_1 \Rightarrow \dots \Rightarrow P_k$$

Ezt egy  $P$ -ből  $Q$ -ba való levezetésnek hívjuk  $W$ -ben. A levezetés hossza  $k$ . A  $\Rightarrow$  egy bináris reláció  $V^*$  halmazon és  $\Rightarrow^*$  egy reflexív, tranzitív lezártja  $\Rightarrow$ -nak. Ha egy  $P$  szóból egy  $Q$  szó egy, vagy több lépésben vezethető le, a  $P \Rightarrow^+ Q$  jelölést alkalmazzuk. Egy  $W$  átírási rendszer átalakítható generatív rendszerré, ha megadunk egy  $A_X \subseteq V^*$  véges, nem üres halmazt, ún. axiómák halmazát.  $W$  hez hozzárendelünk egy  $L_{gW}$  halmazt  $L_{gW} = \{Q | P \Rightarrow^* Q, P \in A_X\}$  ezt a halmazt a  $W$  generatív rendszer által generált nyelvnek hívjuk.

### 4.2. Generatív nyelvtanok

**4.1. Definíció. Generatív nyelvtan** egy olyan  $G = (V_N, V_T, S, H)$  rendezett négyes, ahol:

$V_N$  egy ábécé (nemterminálisok, változók halmaza)

$V_T$  egy ábécé (terminálisok halmaza) úgy, hogy  $V_T \cap V_N = \emptyset$

$S \in V_N$  a mondatzimbólum, kezdőszimbólum, melyből a rendszer axiómái levezethetők

$H : P \rightarrow Q$  alakú átírási szabályok véges halmaza, ahol  $P, Q \in (V_N \cup V_T)^*$  és  $P$  tartalmaz legalább egy nemterminálist

Egy generatív nyelvtan tulajdonképpen egy átírási rendszer, ahol a  $V$  ábécét felbontottuk terminálisokra és nemterminálisokra és kijelöltünk egy kezdőszimbólumot, amiből a rendszer

axiómái levezethetők. Egy adott  $G = (V_N, V_T, S, H)$  generatív nyelvtan által generált nyelvet  $L(G)$  -vel jelöljük.

$$L(G) = \{X \mid X \in V_T^*, S \Rightarrow^* X\}$$

Analitikus nyelvtannak nevezünk egy olyan  $G = (V_N, V_T, S, H)$  rendezett négyest, ahol  $V_N, V_T, S$  rendre nemterminálisok halmaza, terminálisok halmaza, mondatszimbólum,  $H$  pedig  $P \rightarrow Q$  alakú szabályok halmaza, ahol  $P, Q \in (V_N \cup V_T)^*$  és  $Q$  tartalmaz legalább egy nemterminálist. Az analitikus nyelvtan is átírási rendszer.  $G$  analitikus nyelvtan által generált nyelv:

$$L(G) = \{X \mid X \in V_T^*, X \Rightarrow^* S\}$$

Ha  $G$  generatív nyelvtan,  $L(G)$  a  $G$  generatív nyelvtan által generált nyelv, ha  $G$  analitikus nyelvtan,  $L(G)$  a  $G$  analitikus nyelvtan által felismert nyelv.

$G_1$  és  $G_2$  generatív(, vagy analitikus) nyelvtan ekvivalens, ha  $L(G_1) \setminus \{\lambda\} = L(G_2) \setminus \{\lambda\}$ .

### Chomsky féle nyelvosztályok

A nyelveket különböző osztályokba soroljuk attól függően, hogy a generáló nyelvtanjaik levezetési szabályai milyen feltételeknek tesznek eleget. A csoportosítás a következő:

**4.2. Definíció.** Egy  $G = (V_N, V_T, S, H)$  generatív nyelvtan  $i = 0, 1, 2, 3$  típusú, ha az alább megadott  $i$ -edik megszorítások teljesülnek  $H$ -ra (a már korábban definiált megszorítás mellett, mely szerint a szabályok bal oldalán szerepel legalább egy nemterminális):

0:  $H$  szabályaira semmilyen további megszorítás nincs.

1:  $H$  szabályai  $P_1 P P_2 \rightarrow P_1 Q P_2$  alakúak, ahol  $P_1, P_2 \in (V_N \cup V_T)^*$ ,  $P \in V_N$  és  $Q \in (V_N \cup V_T)^+$ , vagy  $S \rightarrow \lambda$  alakú azzal a feltétellel, hogy, ha  $S \rightarrow \lambda$  szerepel a szabályok között, akkor  $S$  nem fordulhat elő egyetlen  $H$  belüli szabály jobb oldalában sem.

2:  $H$  szabályai  $P \rightarrow Q$  alakúak, ahol  $P \in V_N$ ,  $Q \in (V_N \cup V_T)^*$ .

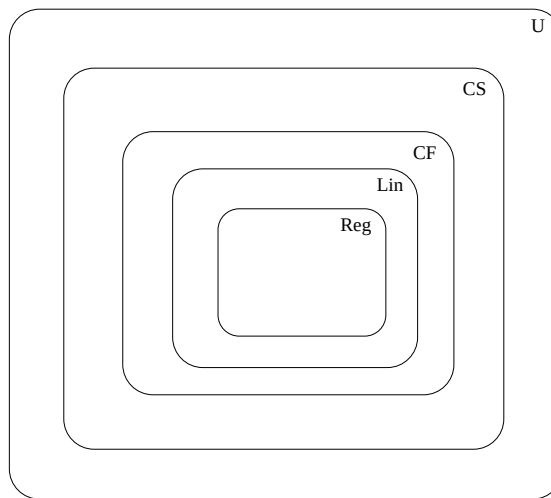
3:  $H$  szabályai  $P \rightarrow xY$ , vagy  $P \rightarrow x$  alakúak, ahol  $P, Y \in V_N$  és  $x \in V_T^*$ .

Az ilyen 0, 1, 2, 3 feltételeknek megfelelő nyelvtanokat rendre mondatszerkezetű, környezetfüggő, környezetfüggetlen, reguláris nyelvtanoknak hívjuk. Továbbá a fentebbi nyelvtanokon kívül bevezetjük a lineáris nyelvtan fogalmát, ahol a  $H$ -beli szabályok  $P \rightarrow aQb$ , vagy  $P \rightarrow a$  alakúak, úgy, hogy  $P, Q \in V_N$ ,  $a, b \in V_T^*$ . Továbbá egy lineáris nyelvtan lehet bal-lineáris, ha csak  $P \rightarrow Yx$  alakú szabályokat, vagy jobb-lineáris, ha csak  $P \rightarrow xY$  alakú szabályokat tartalmaz a  $P \rightarrow x$  alakú szabályok mellett, ahol  $P, Y \in V_N$ ,  $x \in V_T^*$ .

**4.3. Definíció.** Egy adott  $L \subseteq V^*$  nyelvet  $i$  típusúnak mondunk, ha  $G = (V_N, V_T, S, H)$   $i$  típusú nyelvtan generálja. Tehát, ha  $L = L(G)$ .

Az  $i$  típusú nyelvek halmazát  $\mathcal{L}_i$ -vel jelöljük. Így 4 nyelvosztályt különböztetünk meg, ezek a Chomsky-nyelvosztályok. Könnyen látható, hogy bármely nyelv mondatszerkezetű és minden reguláris nyelv egyben környezetfüggetlen is. Továbbá minden környezetfüggetlen nyelvtanhoz megadható vele ekvivalens környezetfüggetlen nyelvtan, mely egyben 1-es típusú is (üresszó lemma). Tehát:

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



1. ábra. Chomsky-hierarchia. U = unrestricted, mondatszerkezetű nyelvek; CS = context-sensitive, környezetfüggetlen nyelvek; CF = context-free, környezetfüggetlen nyelvek; Lin = lineáris nyelvek; Reg = reguláris nyelvek

A dolgozat további részében –ha külön nincs jelölve– a magyar ábécé nyomtatott nagybetűi nemterminális szimbólumokat, a kisbetűk, valamint számjegyek, terminális szimbólumokat jelölnek.

## 5. Reguláris nyelvek

A 3-as típusú grammatikák által generált nyelvek a reguláris nyelvek. Tehát a reguláris nyelvek  $A \rightarrow aB$  és  $A \rightarrow a$  alakú szabályokkal állíthatók elő, ahol  $A, B \in V_N$ ,  $a \in V_T^*$ . Széles körben használják őket lexikális elemzésnél, szövegszerkesztőkben és számos véges állapotú rendszer modellezésére.

### 5.1. Reguláris halmazok és reguláris kifejezések

Reguláris nyelvek megadásának egy módja reguláris kifejezések használata. Adott  $V$  ábécé felett a következő alapvető reguláris kifejezések léteznek:

- $\emptyset$  reguláris kifejezés, mely az üres halmazt jelöli
- $\lambda$  reguláris kifejezés, mely a  $\{\lambda\}$  halmazt jelöli
- minden  $a \in V$  re létezik a reguláris kifejezés, úgy, hogy  $a = \{a\}$
- Ha  $P$  és  $R$  reguláris kifejezések, akkor  $P + R$ ,  $PR$  és  $P^*$  is reguláris kifejezések, melyek rendre a  $P \cup R$ ,  $\{p \cdot r = pr \mid p \in P, r \in R\}$ , és  $\{p^* \mid p \in P\}$  halmazokat jelölik.

Ezekkel a kifejezésekkel és műveletekkel megadott halmazok reguláris nyelvek és minden reguláris nyelvet meg lehet adni reguláris kifejezésekkel (reguláris kifejezések és reguláris nyelvek ekvivalenciájának bizonyítását lsd. például: [6]).

### 5.2. Véges automaták

A véges automata, mint, ahogy a neve is mutatja, olyan absztrakt eszköz, mely csak véges, fix adatokon operál lineáris időben, továbbá nincs belső memóriája. Véges sok belső állapota van, csak az éppen beolvasott szimbólumtól és az adott időpillanatban a belső állapotától függ a működése. Attól függően, hogy a véges automata egy időben egyszerre hány állapotban lehet, beszélünk determinisztikus és nemdeterminisztikus véges automatákról.

**5.1. Definíció.** Determinisztikus véges automata egy  $A = (Q, V, \delta, q_0, F)$  rendezett ötös, ahol:

- $Q$ : belső állapotok egy nem üres, véges halmaza
- $V$ : input szimbólumok véges halmaza
- $\delta : Q \times V \rightarrow Q$ : átmenetfüggvény. Ez mondja meg, hogy valamelyik állapotban egy adott bemenő szimbólumra az automata melyik állapotba megy át.

- $q_0 \in Q$ : a kezdőállapot
- $F \subseteq Q$ : végállapotok halmaza

*Megjegyzés.* A továbbiakban automatán determinisztikus véges automatát értek.

Az automata működése a következő: az  $A = (Q, V, \delta, q_0, F)$  automata a  $q \in Q$  aktuális állapotából az  $a \in V$  szimbólum hatására átmegy  $q' \in Q$  állapotba, ha  $\delta(q, a) = q'$ .

Legyen  $A = (Q, V, \delta, q_0, F)$  egy automata,  $q \in Q$  az automata egy tetszőleges állapota és  $a \in V$  egy tetszőleges szimbólum. Ha  $\delta(q, a)$  nem definiált, azt mondjuk, hogy az automata átmenetfüggvénye a  $q$  állapotban  $a$  bemenő jel hatására nem definiált.

Ha a  $\delta$  átmenetfüggvénnyel szeretnénk reprezentálni az automata működését, csak néhány szimbólumból álló szavak esetén lenne áttekinthető. Vegyünk egy  $A = (Q, V, \delta, q_0, F)$  automatát és legyen  $x \in V^*$ ,  $x = x_1, x_2, \dots, x_n$  egy  $V$  ábécé feletti szó  $A$  bemenete. Ekkor az automata működése formálisan a következő lenne:

$$\delta(\dots \delta(\delta(q_0, x_1), x_2) \dots x_n). \quad (1)$$

Ennek létezik egy egyszerűsített írásmódja, ha kiterjesztjük az átmenetfüggvényt. Jelöljük  $\delta$  kiterjesztését  $\hat{\delta}$ -val, ekkor (1)-et írhatjuk a következőképpen is:

$$\hat{\delta}(q_0, x).$$

Akkor szokás ezt a jelölést használni, ha a köztes állapotok ismeretére nincs szükség, csak arra vagyunk kíváncsiak, hogy az automata az adott állapotból adott szót (vagy rész szót) fel tudja-e dolgozni.

**5.2. Definíció.** Jelölje  $L(A)$  az  $A = (Q, V, \delta, q_0, F)$  automata által elfogadott nyelvet.  $L(A)$  olyan  $w \in V^*$  szavak halmaza, melyeket az automata a kezdőállapotból kiindulva a szó végigolvasásával végállapotba kerül. Formálisan:

$$L(A) = \{w \mid \hat{\delta}(q_0, w) \in F\}$$

**5.3. Definíció.** Nemdeterminisztikus véges automata egy  $A = (Q, V, \delta, q_0, F)$  rendezett ötös, ahol:

- $Q$ : az automata állapotainak véges halmaza
- $V$ : az automata által elfogadott szavak ábécéje
- $\delta$ : egy  $Q \times V \rightarrow 2^Q$  leképezés, átmenetfüggvény
- $q_0 \in Q$ : kezdőállapot

- $F$ : az elfogadó állapotok véges halmaza

Láthatjuk, hogy a determinisztikus véges automatától csak az átmenetfüggvényében tér el.

A nemdeterminisztikus véges automata átmenetfüggvénye megengedi, hogy az automata több állapotban is legyen egyszerre. Ez úgy képzelhető el, hogy az adott automata egyszerre több példányban is létezhet.

A nemdeterminisztikus véges automata működése a következőként írható le:

Legyen  $A = (Q, V, q_0, \delta, F)$  egy nemdeterminisztikus véges automata. Jelöljön  $a \in V$  egy szimbólumot és  $q \in Q$  az automata egy állapotát. Az  $A$  automata a  $q$  állapotból az  $a$  bemenő jel hatására átmegy a  $q_1, q_2, \dots, q_n \in Q$  ( $n \geq 0$ ) állapotok valamelyikébe, ha  $\delta(q, a) = \{q_1, q_2, \dots, q_n\}$ .

Egy  $A = (Q, V, q_0, \delta, F)$  nemdeterminisztikus véges automata átmenetfüggvénye teljesen specifikált, ha:

$$\text{minden } q \in Q \text{-ra, minden } a \in V \text{-re : } \delta(q, a) \neq \emptyset$$

A determinisztikus véges automatáknál definiált  $\hat{\delta}$  kiterjesztett átmenetfüggvény a nemdeterminisztikus automatáknál is értelmezhető. Adott egy  $A = (Q, V, q_0, \delta, F)$  nemdeterminisztikus véges automata és egy  $x \in V^*$ ,  $x = x_0, x_1, \dots, x_n$  szó. Az  $A$  automata működése formálisan:

$$P_0 = \delta(q_0, x_0), P_1 = \cup_{q_1 \in P_0} \delta(q_1, x_1), \dots, P_n = \cup_{q_n \in P_{n-1}} \delta(q_n, x_n).$$

A  $\hat{\delta}$  kiterjesztett átmenetfüggvény használatával:

$$\hat{\delta}(q_0, x)$$

**5.1. Tétel.** *A nemdeterminisztikus véges automaták által felismert nyelvek osztálya megegyezik a determinisztikus véges automaták által felismert nyelvek osztályával.*

*Megjegyzés.* A nemdeterminisztikus automatákat NDA, a determinisztikus automatákat DA rövidítésekkel jelölöm.

*Bizonyítás.* (1) Az  $\mathcal{L}_{DA} \subseteq \mathcal{L}_{NDA}$  tartalmazás könnyen belátható, mivel a determinisztikus automata a nemdeterminisztikus automata egy speciális esetének tekinthető, ahol

$$0 \leq |\delta(q, a)| \leq 1 \quad \forall q \in Q, \quad \forall a \in V.$$

(2)  $\mathcal{L}_{NDA} \subseteq \mathcal{L}_{DA}$  igazolásához megadunk egy algoritmust, mellyel egy NDA -hoz mindig konstruálható vele ekvivalens DA.

Legyen  $A = (Q, V, \delta, q_0, F)$  egy NDA és  $A' = (Q', V, \delta', q'_0, F')$  egy DA, ahol

- $Q' = 2^Q$

- $F' = \{S \in Q' \mid S \cap F \neq \emptyset\}$
- $\delta' : Q' \times V \rightarrow Q'$  úgy, hogy  $\delta'(S, a) = P$ , ahol  $P = \cup_{q \in S} \delta(q, a)$   
 $\forall S \in Q', \forall a \in V$  -re
- $q'_0 = \{q_0\} \in Q'$

Vegyük egy  $w \in L(A)$  szót.

$$w = w_0 w_1 \dots w_n, \quad w_i \in V \quad \text{minden } i = 0, 1, \dots, n \text{ -re.}$$

Legyen  $w = \lambda$  és  $\lambda \in L(A)$ , ami csak akkor lehetséges, ha  $q_0 \in F$ . Ekkor  $F'$  definíciója szerint  $\{q_0\} \in F'$  is teljesül, tehát  $\lambda \in L(A')$ .

Ha  $w \neq \lambda$ , akkor  $w$ -t  $A$  felismeri  $q_0$ -ból kiindulva véges sok lépésben:

$$\delta(q_0, w_0) = P_0, \cup_{s \in P_0} \delta(s, w_1) = P_1, \dots, \cup_{s \in P_{n-1}} \delta(s, w_n) = P_n, \quad P_n \cap F \neq \emptyset.$$

A definíció szerint  $P_i \in 2^Q = Q'$  minden  $i = 0, \dots, n$ -re, tehát  $A'$  automata állapotai között szerepel az összes  $P_i$ , valamint  $\{q_0\}$  is, ami  $q'_0$ -vel egyezik meg. A  $\delta'$  függvény definíciója értelmében

$$\delta'(q'_0, w_0) = P_0 \quad \text{és} \quad \delta'(P_i, w_{i+1}) = \cup_{s \in P_i} \delta(s, w_{i+1}) = P_{i+1}$$

minden  $i = 0, \dots, n-1$ -re, valamint  $P_n \in F'$ . Látható, hogy  $A'$  automata működése determinisztikus és  $w$  szót felismeri véges sok lépésben (pontosabban  $n = |w|$  lépésben). Továbbá a  $\delta'$  függvény és  $F'$  halmaz definíciója miatt  $A'$  akkor és csak akkor ismer fel egy  $v \in V^*$  szót, ha azt  $A$  is felismeri. Ezzel a tételt bizonyítottunk tekinthetjük.  $\square$

A következőkben megmutatjuk, hogy a véges automaták által felismert nyelvek osztálya megegyezik a regulás nyelvek osztályával. Ehhez a reguláris nyelveken végezhető néhány átalakítást meg kell ismerni.

**5.2. Tétel.** Minden  $G = (V_N, V_T, S, H)$  reguláris grammatikához megadható vele ekvivalens nyelvet generáló  $G' = (V_N, V_T, S, H')$  reguláris nyelvtan, ahol  $H'$  nem tartalmaz  $A \rightarrow B$ ,  $A, B \in V_N$  alakú szabályokat.

*Bizonyítás.* A bizonyítás triviális. Először  $H' = \{A \rightarrow \beta \mid \beta \notin V_N\}$ , majd minden  $A \rightarrow B$  alakú szabály helyett felveszünk  $H'$ -be  $A \rightarrow \alpha$  szabályokat, ahol  $\alpha$  a  $B$  baloldali  $H$  belső szabályok jobb oldala. Az  $A \rightarrow B$  szabályt pedig nem vesszük fel  $H'$ -be. Így véges sok lépésben az összes ilyen szabályt kiküszöbölhetjük.  $\square$

**5.3. Tétel.** Minden  $G = (V_N, V_T, S, H)$  reguláris grammatikához megadható vele ekvivalens nyelvet generáló  $G' = (V_N, V_T, S, H')$  reguláris nyelvtan, melyben  $\lambda$  csak  $S \rightarrow \lambda$  alakú szabály jobb oldalán lép fel.

*Bizonyítás.* Először 5.2 tétel bizonyításában megadott módon létrehozzuk  $G'$  nyelvtant, majd minden  $A \rightarrow aB$  alakú  $H'$  -beli szabályra, ha létezik  $B \rightarrow \lambda$  szabály, vegyük fel  $H'$  -be az  $A \rightarrow a$  szabályt, a  $B \rightarrow \lambda$  szabályt pedig töröljük  $H'$  -ből. Így véges sok lépés után  $\lambda$  csak  $S \rightarrow \lambda$  szabály jobb oldalán fordulhat elő.  $\square$

Ezen átalakítások után, ha valamely  $B$  nemterminálisra  $S \Rightarrow \alpha B \beta$  nem teljesül, akkor a  $B$  baloldalú szabályokat eltávolíthatjuk. A dolgozat további részében reguláris nyelvtan alatt az 5.3 és 5.2 tételekben megadott alakú nyelvtant értek.

**5.4. Tétel.** Minden  $G = (V_N, V_T, S, H)$  reguláris grammatikához megadható vele ekvivalens nyelvet generáló  $G' = (V'_N, V_T, S, H')$  reguláris nyelvtan, melynek összes szabálya  $A \rightarrow aB$ , és  $A \rightarrow a$  alakú, ahol  $A, B \in V_N$ , és  $a \in V_T$ .

*Bizonyítás.* Legyen  $G = (V_N, V_T, S, H)$  egy reguláris nyelvtan.

Adjuk meg a  $G' = (V'_N, V_T, S, H')$  reguláris nyelvtant a következő szabályok alkalmazásával:

1. legyen  $V'_N = V_N$  és  $H' = H$
2. minden  $A \rightarrow a\alpha \in H'$  szabályra, ahol  $|\alpha| > 1$ , bővítsük  $H'$  -t  $A \rightarrow aB$  és  $B \rightarrow \alpha$  szabályokkal, ahol  $B$  olyan nemterminális, ami nem szerepel  $V'_N$  -ben. Majd vegyük fel  $B$  -t  $V'_N$  -be és töröljük az  $A \rightarrow a\alpha$  szabályt  $H'$  -ből. (Látható, hogy egy  $aB \in (V'_N \cup V_T)^*$  szóra, ha alkalmazzuk a  $B \rightarrow \alpha$  szabályt, az  $a\alpha$  szót kapjuk, tehát a nyelvtan által generált nyelv nem változik.)
3. ismételjük meg a 2. pontot, amíg vannak  $A \rightarrow a\alpha$  alakú szabályok  $H'$  -ben, melyekre  $|\alpha| > 1$ .

Ezzel előállítottuk a  $G' = (V'_N, V_T, S, H')$  reguláris nyelvtant, melynek szabályai megfelelő alakúak és  $L(G') = L(G)$ .  $\square$

**5.5. Tétel.** A véges automaták által felismert nyelvek osztálya reguláris.

*Bizonyítás.* Megmutatjuk, hogy a reguláris nyelvtanok által generált nyelvek osztálya megegyezik a véges automaták által felismert nyelvek osztályával. Elég csak azokkal a reguláris nyelvtanokkal foglalkozni, melyeknek a  $H$  -beli szabályai  $A \rightarrow a$  és  $A \rightarrow aB$  alakúak, mivel az 5.4 tétel értelmében bármely reguláris nyelvtanhoz megadható olyan vele ekvivalens nyelvtan, amelyre teljesülnek ezek a feltételek. Két algoritmust ismertetek. Az első algoritmus alapján reguláris nyelvtanhoz nemdeterminisztikus véges automatát lehet konstruálni, a második algoritmus pedig azt mondja meg, hogy nemdeterminisztikus véges automatához hogyan lehet előállítani reguláris nyelvtant.

Első eset, reguláris grammatika  $\longrightarrow$  nemdeterminisztikus véges automata:

Legyen  $G = (V_N, V_T, S, H)$  egy reguláris grammatika, melynek szabályaira teljesülnek az előbb megfogalmazott kikötések. Adjuk meg az  $A = (Q, V, \delta, q_0, F)$  automatát a következő módon:

1. Legyen  $Q = V_N \cup \{Z\}$ , ahol  $Z \notin (V_N \cup V_T)$ ,  $V = V_T$ ,  $q_0 = \{S\}$  és  $F = Z$ . Ha  $S \rightarrow \lambda \in H$ , akkor az  $S \in Q$  állapotot is vegyük fel a végállapotok közé.
2. És minden  $A \rightarrow \alpha \in H$ -ra, ha  $\alpha \in V_T$ , akkor legyen  $\delta(A, \alpha) = Z$ , egyébként  $\alpha aB$  alakú, ahol  $a \in V_T$ ,  $B \in V_N$ , ekkor legyen  $\delta(A, a) = B$ .

Könnyen belátható, hogy az automata lépései a helyettesítési szabályoknak megfelelően történnek. Legyen

$$S \Rightarrow x_0 X_0 \Rightarrow x_0 x_1 X_1 \Rightarrow^* x_0 x_1 \dots x_{n-1} X_{n-1} \Rightarrow x_0 x_1 \dots x_n$$

egy  $G$ -beli levezetés, ahol  $x_0, x_1, \dots, x_n \in V_T$  és  $X_0, X_1, \dots, X_{n-1} \in V_N$ . Látható, hogy minden  $\alpha x_i X_i$  szóra az  $X_i \rightarrow x_{i+1} X_{i+1} \in H$  szabályt alkalmaztuk, kivéve az utolsó esetet, amikor  $X_{n-1} \rightarrow x_n \in H$  szabály szerint jártunk el. Ugyanez a levezetés az automatában a következő:

$$\delta(\delta(\delta(S, x_0), x_1), x_{n-1}), x_n) \cap F \neq \emptyset.$$

A nemdeterminisztikus véges automata  $\longrightarrow$  reguláris nyelvtan esetben  $V_N = Q$ ,  $V_T = V$ ,  $S = q_0$  és a  $H$ -beli szabályok a következő módon állnak elő: minden  $\delta(q, a) = q'$ -re legyen  $q \rightarrow aq' \in H$ , valamint minden  $q \in F$ -re legyen  $q \rightarrow \lambda \in H$ .  $\square$

### 5.3. Véges automaták megadása

Véges automatákat megadhatunk átmenet-diagrammal, mely egy irányított, címkézett éleket tartalmazó gráf. A gráf csomópontjai az automata állapotai, az élek pedig az átmeneteket ábrázolják. Egy  $A = (Q, V, \delta, q_0, F)$  nemdeterminisztikus véges automata bármely  $q, s \in Q$  állapotaira és  $a \in V$  szimbólumra, ha  $s \in \delta(q, a)$ , akkor a  $q$  állapottal jelölt csomópontból  $a$ -val címkézett irányított él mutat  $s$  csomópontba. Ha egy, vagy több  $a$ -tól különböző  $b, c, \dots$  szimbólumra is teljesül, hogy  $s \in \delta(q, b)$ ,  $s \in \delta(q, c), \dots$ , akkor nem szükséges új élt felvenni, a meglévőnek a címkéjét meg lehet változtatni  $a, b, c, \dots$ -re. A kezdőállapotot a „semmiből,” bele mutató  $\rightarrow$  különbözteti meg a többi csomóponttól, a végállapotokat pedig dupla körrel jelöljük.

Vagy megadhatunk véges automatát Cayley táblázattal is, melynek első sorának első oszlopa az automata nevét, az első oszlop többi része pedig az automata állapotait tartalmazza.

Az automata első sorában szerepelnek az input ábécé szimbólumai. A táblázat  $i, j$  -edig eleme pedig azt mondja meg, hogy az  $i$  -edik sorban lévő állapotból a  $j$  -edik oszlopban lévő szimbólumra mely állapot(ok)ba megy át az automata. A táblázatban a végállapotokat  $*$  -al jelöljük, a kezdőállapotot pedig  $a \rightarrow$  különbözteti meg a többitől.

## 5.4. Reguláris nyelvek tulajdonságai

**5.6. Tétel.** *A reguláris nyelvek osztálya zárt az unióképzés, konkatenáció és  $*$  műveletekre.*

*Bizonyítás.* Legyenek  $L_1, L_2$  reguláris nyelvek, melyeket rendre

$$G_1 = (V_{N1}, V_{T1}, S_1, H_1)$$

$$G_2 = (V_{N2}, V_{T2}, S_2, H_2)$$

reguláris nyelvtanok generálnak. Feltehetjük, hogy  $V_{N1}, V_{N2}$  diszjunkt halmazok. Ha mégsem, akkor a változók átnevezése nem változtat a grammatika által generált nyelven. Legyen

$$G = (V_N, V_T, S, H)$$

ahol  $V_N = V_{N1} \cup V_{N2}, V_T = V_{T1} \cup V_{T2}$  és:

- $L_1 \cup L_2$  esetben  $H = H_1 \cup H_2 \cup \{S \rightarrow S_1, S \rightarrow S_2\}$ .
- $L_1 L_2$  esetben  $H = H_1 \cup H_2 \cup \{S \rightarrow S_1 S_2\}$
- $L_1^*$  esetben  $H = H_1 \cup \{S \rightarrow S_1 S, S \rightarrow \lambda\}$

A  $G$  grammatika mindhárom esetben reguláris. □

**5.7. Tétel.** *A reguláris nyelvek osztálya zárt a komplementerképzésre.*

*Bizonyítás.* Legyen  $V$  egy ábécé és  $L \subseteq V^*$  nyelv reguláris. Ekkor a 4.1 és 4.2 -es tétel következtében létezik olyan  $A = (Q, V', \delta, q_0, F)$  determinisztikus véges automata, amelyre  $L(A) = L$ . A  $V'$  ábécé azokat a szimbólumokat tartalmazza, melyekből  $L$  szavai állnak. Legyen  $A_c = (Q, V, \delta, q_0, Q - F)$ . Tehát  $A_c$  végállapotai azok az állapotok lesznek, melyek  $A$  -ban nem végállapotok. Ekkor könnyen látható, hogy  $L(A_c) = V^* - L$ . □

A 2. fejezetben bevezettük a relációk fogalmát. Egy reláció ekvivalencia reláció, ha reflexív, szimmetrikus és tranzitív. Most bevezetjük a kongruencia fogalmát.

**5.4. Definíció.** Egy  $\sim \subseteq V^* \times V^*$  reláció kongruencia reláció, ha  $\sim$  ekvivalencia reláció, továbbá minden  $u, v, x, y \in V^*$  -ra  $u \sim v$ , akkor és csak akkor áll fenn, ha  $xuy \sim xvy$  is teljesül.

A  $\sim$  reláció jobb-kongruencia, ha minden  $u, v, x \in V^*$  -ra  $u \sim v$  akkor és csak akkor, ha  $ux \sim vx$ .

A kongruencia reláció nyelvekre megfogalmazott változata a következő.

**5.5. Definíció.** Adott egy  $V^*$  ábécé és egy  $L \subseteq V^*$  nyelv. A  $\sim$  reláció  $L$  feletti kongruencia reláció, ha teljesül a következő feltétel:

$$\text{minden } u, v \in L \text{ -re } u \sim v \Rightarrow \text{minden } x \in V^* \text{ -ra } ux \in L \Leftrightarrow vx \in L.$$

Tehát a kongruencia reláció értelmében  $\hat{\delta}(q_0, u) = \hat{\delta}(q_0, v)$ .

## 6. Környezetfüggetlen nyelvek

A környezetfüggetlen nyelvek is nagy gyakorlati jelentőséggel bírnak főleg rogramozási nyelvek definiálásánál, fordítóprogramokban, és egyéb szöveges adatok feldolgozásánál. Elsősorban a természetes nyelvek jellemzésére akarták használni a környezetfüggetlen nyelveket, de a szemantikai információk modellezésére nem alkalmasak.

### 6.1. Veremautomaták

Láthattuk, hogy a véges automaták meglehetősen kötött eszközök a nyelvek feldolgozásánál. Egyik legnagyobb hátrányuk az, hogy a már feldolgozott szimbólumokat nem tárolják. A veremautomaták nyújtanak erre megoldást. A veremautomaták olyan véges automatáknak tekinthetők, melyek rendelkeznek egy LIFO elérésű, végtelen méretű memóriával, egy veremmel. A vermen két művelet hajtható végre: tetszőleges szimbólumsorozat verem tetejére írása (PUSH), valamint egy szimbólum kiolvasása és törlése a verem tetejéről (POP). Egy veremautomata definíciója formálisan a következő:

**6.1. Definíció.** Egy  $A = (Q, V, \Gamma, \delta, q_0, z_0, F)$  hetes veremautomata, ahol:

- $Q$  a belső állapotok véges halmaza
- $V$  input-ábécé
- $\Gamma$  a verem-ábécé ( $\Gamma \cap V = \emptyset$ )
- $\delta$  egy  $Q \times (V \cup \{\lambda\}) \times \Gamma \rightarrow 2^{Q \times \Gamma^*}$  alakú átmenetfüggvény.
- $q_0$  a kezdőállapot
- $z_0 \in \Gamma$  a verem alja jel
- $F$  a végállapotok halmaza

Veremautomatáknál is definiálunk kombinációkat  $(w, q, \alpha)$  alakban, ahol  $w \in V^*$  a be-menő szónak a még el nem olvasott része,  $q \in Q$  az aktuális állapot és  $\alpha \in \Gamma^*$  a verem aktuális tartalma. Egy  $(xw, q, a\alpha)$  konfigurációból elérhető a  $(w, q', b\alpha)$  konfiguráció (jelöléssel  $(xw, q, a\alpha) \vdash (w, q', b\alpha)$ ), ha  $(q', b) \in \delta(q, x, a)$ , ahol  $x \in (V \cup \{\lambda\})$ ,  $w \in V^*$ ,  $q, q' \in Q$ ,  $a \in \Gamma, \alpha, b \in \Gamma^*$ . A  $\vdash$  jelölésnek  $\vdash^*$  a reflexív, tranzitív lezártja. Ekkor a veremautomata által elfogadott nyelv:

$$L(A) = \{w \in V^* \mid (w, q_0, z_0) \vdash^* (\lambda, q_v, \lambda), \quad q_v \in F\}.$$

Veremautomatáknak van olyan változata is, amely akkor fogad el egy szót, ha végállapotba lép, de ez ekvivalens a végállapotban üres veremmel elfogadó automatával. Az előbbi automatahoz csak fel kell venni olyan szabályokat, hogy, ha végállapotba ér, akkor kiüríti a vermet. Attól függően, hogy a  $\delta$  függvény adott  $(q, x, a)$  hármásra csak egy, vagy több  $(q', b)$  párost is ad eredményül, beszélhetünk determinisztikus és nemdeterminisztikus veremautomatákról. Nyilvánvaló, hogy a nemdeterminisztikus veremautomaták speciális esetei a determinisztikus automaták.

## 6.2. Környezetfüggetlen nyelvek tulajdonságai, átalakításai

A Chomsky-hierarchia szerint  $G = (V_N, V_T, S, H)$  nyelvtan környezetfüggetlen, ha a  $H$  -beli szabályok  $P \rightarrow Q$  alakúak, ahol  $P \in V_N$  nemterminális szimbólum,  $Q \in (V_N \cup V_T)^*$  pedig nemterminális és terminális szimbólumok tetszőleges sorozata. A következőkben olyan átalakításokról lesz szó, melyek segítségével adott  $G$  nyelvtanokhoz konstruálhatunk vele ekvivalens  $G'$  nyelvtant bizonyos vizsgálatok megkönnyítésére.

**6.1. Lemma.** *Adott  $G = (V_N, V_T, S, H)$  környezetfüggetlen nyelvtanhoz, melyre igaz, hogy  $L(G) \neq \emptyset$ , megadható olyan ekvivalens  $G' = (V'_N, V_T, S, H')$  környezetfüggetlen nyelvtan, melyben minden  $X \in V'_N$  nemterminális szimbólumhoz létezik  $w \in V_T^*$  szó úgy, hogy  $X \Rightarrow_{G'}^* w$  teljesül. Ezek a nemterminális szimbólumok az ún. hasznos nemterminális szimbólumok.*

*Megjegyzés.* Azokat a nemterminálisokat eltávolíthatjuk, melyekből tetszőleges sok lépésben nem vezethető le terminálisok sorozata.

*Bizonyítás.* Legyen  $G = (V_N, V_T, S, H)$  egy környezetfüggetlen nyelvtan, melyre teljesül, hogy  $L(G) \neq \emptyset$ . Állítsuk elő  $U$  halmazt a következő algoritmussal:

1.  $U_0 = \cup_{a \in V_T} \{X \mid X \rightarrow a \in H, \quad X \in V_N, \quad a \in V_T^*\}$
2.  $i = 1$
3.  $U_i = \cup_{X \in U_{i-1}} \{Y \mid Y \rightarrow \alpha X \beta \in H, \quad Y, X \in V_N, \quad \alpha, \beta \in (U_{i-1} \cup V_T)^*\}$
4. Ha  $U_i = U_{i-1}$ , akkor végeztünk és legyen  $U = U_i$ , különben  $i = i + 1$  és menjünk a 3 -as pontra.

Tehát  $U$  tartalmazza azokat a nemterminálisokat, melyekből tetszőleges sok lépésben elérhető nemterminálisokból álló sorozat. Most adjuk meg a  $G' = (V'_N, V_T, S, H')$  nyelvtant úgy, hogy  $V'_N = U$  és  $H' = \{X \rightarrow \alpha \in H \mid X \in V'_N, \alpha \in (V'_N \cup V_T)^*\}$ . Tehát  $H'$  -be csak azok a  $H$  -beli szabályok kerülnek bele, melyek alapján bármely  $X \in V'_N$  hasznos szimbólum. Ez következik

az  $U$  meghatározására használt algoritmusból. Nyilvánvaló, hogy  $G'$  ugyanazt a nyelvet fogja generálni, mint  $G$ .  $\square$

**6.2. Definíció.** Egy  $G = (V_N, V_T, S, H)$  környezetfüggetlen grammatika Chomsky normálformában van, ha szabályai  $P \rightarrow XY$  és  $P \rightarrow x$  alakúak, ahol  $P, X, Y \in V_N$  -beli, nemterminális és  $x \in V_T$  -beli, terminális szimbólumok.

**6.2. Lemma** (Üres-szó lemma). *Minden  $G = (V_N, V_T, S, H)$  környezetfüggetlen nyelvtanhoz megadható vele ekvivalens  $G' = (V'_N, V_T, S', H')$  környezetfüggetlen nyelvtan, melyben  $\lambda$  legfeljebb csak  $S' \rightarrow \lambda$  szabály jobb oldalán lép fel.*

*Bizonyítás.* Csak vázlatosan. Legyen  $G = (V_N, V_T, S, H)$  egy környezetfüggetlen nyelvtan. Vizsgáljuk azokat az  $A \in V_N$  nemterminálisokat, melyekre  $A \Rightarrow^* \lambda$  fennáll. Minden ilyen  $A$  nemterminálisat vegyünk fel egy  $U$  halmazba.  $H'$  először azokból a szabályokból áll, melyek jobb oldalán semmilyen  $U$  -beli nemterminális nem lép fel. A többi  $H$  -beli szabályra vegyünk fel  $H'$  -be olyan új szabályokat, hogy az  $U$  -beli nemterminálisokat minden lehetséges módon elhagyjuk a  $H$  -beli szabályok jobb oldaláról. Az  $A \rightarrow \lambda$  alakú szabályokat ne vegyük fel  $H'$  -be. Ha  $S$  szerepel  $U$  -ban, akkor vegyünk fel egy  $S' \notin V_N$  szimbólumot  $V'_N$  -be és legyen  $S'$  az új kezdőszimbólum, továbbá vegyük fel az  $S' \rightarrow \lambda$  és  $S' \rightarrow S$  szabályokat  $H'$  -be. Ha  $S$  nem szerepel  $U$  -ban, akkor ne vezessünk be új kezdőszimbólumot. Ezzel végeztünk  $G'$  előállításával.  $\square$

**6.3. Definíció.** Adott  $G = (V_N, V_T, S, H)$  környezetfüggetlen grammatikában az  $A \rightarrow B \in H$  szabályokat, ahol  $A, B \in V_N$ , láncszabályoknak nevezzük. Azt a  $G$  nyelvtant, amely nem tartalmaz láncszabályt, láncszabály-mentes nyelvtannak hívjuk.

**6.3. Tétel.** *Minden  $G$  2-es típusú nyelvtanhoz megadható vele ekvivalens  $G'$  2-es típusú láncszabály-mentes nyelvtan.*

*Bizonyítás.* Legyen  $G = (V_N, V_T, S, H)$  2-es típusú nyelvtan és  $H$  tartalmaz legalább egy láncszabályt. Állítsuk elő az  $U$  halmazt:

1. legyen  $U_0 = \{A \in V_N \mid A \rightarrow B \in H\}$  és  $i = 1$
2.  $U_i = \{B \in V_N \mid B \Rightarrow^* C \text{ valamely } C \in U_{i-1} \text{-re}\}$
3. Ha  $U_i = U_{i-1}$ , akkor végeztünk és  $U = U_i$ , különben  $i = i + 1$  és menjünk a 2. pontra.

Minden  $U$  -beli  $A$  nemterminálisra állítsuk elő az  $N_A = \{B \in H \mid A \Rightarrow^* B\}$  halmazt. Most konstruáljuk meg a  $G' = (V_N, V_T, S, H')$  nyelvtant úgy, hogy  $H'$  -be bekerül az összes olyan

$H$  -beli szabály, ami nem láncszabály, majd minden  $A \in U$  -ra és minden  $B \in N_A$  -ra legyen  $A \rightarrow \alpha \in H'$  minden olyan  $B \rightarrow \alpha \in H$  -beli szabályra, ami nem láncszabály. Ezzel a láncszabályokat kiküszöböltük.

Könnyen belátható, hogy  $L(G') = L(G)$ . Tegyük fel, hogy  $A \rightarrow B \in H$  és  $B \rightarrow \alpha \in H$ ,  $A, B \in V_N$ ,  $\alpha \in (V_N \cup V_T)^*$  továbbá valamilyen  $S \Rightarrow^* \beta A \gamma \Rightarrow^* \beta \alpha \gamma$ ,  $\beta, \gamma \in (V_N \cup V_T)^*$  levezetést vizsgálunk. Ez  $G$  -ben először az  $A \rightarrow B$ , majd a  $B \rightarrow \alpha$  szabályok alkalmazásával állhat elő, de  $G'$  -ben is előállítható, mert  $A \rightarrow \alpha \in H'$  és ezt a szabályt alkalmazva is  $\beta \alpha \gamma$  szimbólumsorozathoz jutunk.  $\square$

**6.4. Tétel.** Minden  $G = (V_N, V_T, S, H)$  környezetfüggetlen nyelvtanhoz megadható vele ekvivalens  $G' = (V'_N, V_T, S, H')$  környezetfüggetlen nyelvtan, amely Chomsky-normálformában van.

*Bizonyítás.* Legyen  $G = (V_N, V_T, S, H)$  egy környezetfüggetlen grammatika. Az üres-szó lemma értelmében létezik minden  $G$  nyelvtanhoz vele ekvivalens  $G'$  nyelvtan, melynek szabályaiban  $\lambda$  csak  $S' \rightarrow \lambda$  szabály jobb oldalán fordul elő, ezért feltételezzük, hogy a nyelvtan már ilyen alakú. Továbbá két nyelvtan ekvivalens, ha az általuk generált nyelvek csak  $\lambda$  -ban különböznek egymástól, ezért az  $S \rightarrow \lambda$  szabályt el is hagyhatjuk.

Tehát már csak  $A \rightarrow a$  és  $A \rightarrow \alpha$  alakú szabályaink vannak, ahol  $A \in V_N$ ,  $a \in V_T^+$  és  $\alpha \in (V_N \cup V_T)^+$ . Feltételezzük, hogy  $G$  nem tartalmaz láncszabályokat.

Ekkor megkonstruálhatjuk a  $G' = (V'_N, V_T, S, H')$  nyelvtant. Első lépésként legyen  $V'_N = V_N$  és  $H'$  -be kerüljön az összes  $A \rightarrow a \in H$  és  $A \rightarrow BC \in H$  szabály, ahol  $|a| = 1$  és  $B, C \in V_N$ , ezekkel a szabályokkal a későbbiekben nem foglalkozunk.

Minden  $A \rightarrow \alpha a \beta \in H$  szabályra vezessünk be egy új nemterminálist,  $A_1$  -et  $V'_N$  -be, írjuk át a szabályban  $a$  -t  $A_1$  -re, majd legyen  $A_1 \rightarrow a \in H'$ . Ezt addig végezzük, amíg van olyan szabály, melynek jobb oldalán áll terminális szimbólum.

Ekkor már csak  $A \rightarrow A_1 A_2 \dots A_n$  ( $n \geq 2$ ) alakú szabályaink maradtak. Az  $A \rightarrow A_1 A_2$  szabályokat felvehetjük  $H'$  -be. A többi  $A \rightarrow A_1 A_2 \dots A_n$  szabályra a következőképpen járjunk el: vezessünk be  $A'$  nemterminálist, amely nem szerepel  $V'_N$  -ben és legyen  $A' \rightarrow A_2 \dots A_n$  és vegyük fel  $A \rightarrow A_1 A'$  -t  $H'$  -be, majd járjunk el hasonlóan az  $A' \rightarrow A_2 \dots A_n$  szabálynál is. Végül sok lépés után az összes szabályt átalakítottuk és  $L(G) = L(G')$  is teljesül.  $\square$

**6.5. Tétel.** Bármely környezetfüggetlen nyelv felismerhető veremautomatával.

$\square$

**6.6. Tétel.** Tetszőleges veremautomatával felismerhető nyelv környezetfüggetlen.

$\square$

### 6.3. Veremautomaták megadása

Veremautomatát megadhatunk átmenet-diagrammal, vagy cayley-táblázattal. A veremautomata átmenet-diagrammja hasonló a véges automatáéhoz, azzal az eltéréssel, hogy a  $q, q'$  állapotok közötti átmeneteket egy  $a, z|z'$  hármassal jellemezzük, ahol  $a \in (V \cup \{\lambda\})$  a bemenetről olvasott szimbólum,  $z \in (\Gamma \cup \{\lambda\})$  a verem tetején lévő szimbólum,  $z' \in \Gamma^*$  pedig a verem tetejére kerülő új szimbólumsorozat. Megengedett, hogy  $z = z'$ . Továbbá, ha minden átmenet  $a, \lambda|\lambda$  alakú, akkor a veremautomata pontosan egy véges automatát szimulál.

A veremautomata Cayley-táblázatának sorai pedig a következőképpen írják le az automata működését: a sor első eleme az automata aktuális állapotát jelöli, a második elem az olvasott szimbólum, a harmadik a verem tetején lévő szimbólum, a negyedik elem azt mondja meg, hogy az automata mely állapotba megy át, a sor ötödik eleme pedig azt írja le, hogy milyen műveletet hajtson végre az automata: pop – a verem tetején lévő szimbólum törlése, push( $z$ ) – a verem tetején lévő szimbólum helyettesítése  $z$ -vel, ahol  $z \in \Gamma^+$ .

### 6.4. Derivációs fák, levezetések

**6.4. Definíció.** Legyen  $G = (V_N, V_T, S, H)$  egy környezetfüggetlen nyelvtan. Egy  $w \in L(G)$  szó legbaloldalibb (legjobboldalibb) levezetésén azt az

$$x_0 \Rightarrow x_1 \Rightarrow x_2 \Rightarrow \cdots \Rightarrow x_n \quad (n \geq 0)$$

sorozatot értjük, ahol  $x_i \in (V_N \cup V_T)^*$  minden  $i = 0, \dots, n$ -re,  $x_0 = S$ ,  $x_n = w$  és minden  $x_{i+1}$  ( $i = 0, 1, \dots, n-1$ ) úgy jön létre, hogy  $x_i$ -nek a legbaloldalibb (legjobboldalibb) nemterminálisát helyettesítjük.

Legyen  $G = (V_N, V_T, S, H)$  egy környezetfüggetlen generatív grammatika. Egy  $w \in L(G)$ ,  $w = w_1 w_2 \dots w_n$  ( $w_i \in V_T$ ,  $i = 1, 2, \dots, n$ ) szó derivációs gráfján (vagy pontosabban derivációs fáján) azt a fa struktúrát értjük, melynek gyökere  $S$  és a levélelemei balról jobbra olvasva sorra  $w_1, w_2, \dots, w_n$  nemterminálisokból állnak.

Általánosabban egy  $G = (V_N, V_T, S, H)$  nyelvtanhoz tartozó  $T$  derivációs fa a következőképpen néz ki:

1.  $T$  gyökere  $S$
2. bármely  $x \in T$  elemre, ha  $x$  nem levélelem, akkor  $x \in V_N$ , tehát a köztes elemek  $V_N$ -beli nemterminálisokat reprezentálnak
3. a levélelemek pedig  $V_T$ -beli terminális szimbólumok

Minden  $A \in V_N$  csomópontra  $A$  gyermekei valamely  $A \rightarrow \alpha \in H$  alakú szabály alapján állnak elő. Legyen  $\alpha = \alpha_1 \alpha_2 \dots \alpha_n$ , ahol  $|\alpha_i| = 1$  minden  $i = 1, 2, \dots, n$ -re. Ekkor az  $A$  csomópont gyermekei balról jobbra haladva  $\alpha_1, \alpha_2, \dots, \alpha_n$  elemek lesznek.

Egy  $w$  szóhoz tartozó levezetési fát  $T_w$ -vel jelöljük.

**6.7. Tétel.** Legyen  $G = (V_N, V_T, S, H)$  egy Chomsky normálformában lévő környezetfüggetlen grammatika és  $S \Rightarrow^* w$  egy levezetés, ahol  $w \in V_T^*$ . A  $T_w$  fa mélysége  $n$ . Ekkor  $w$  hossza legfeljebb  $2^{n-1}$ .

*Bizonyítás.* Könnyen belátható. Tudjuk, hogy a szabályok alakja  $A \rightarrow a$ , vagy  $A \rightarrow BC$  lehet. Ha  $|w| = 1$ , akkor a hozzá tartozó fa magassága pontosan 1, mivel  $S \Rightarrow w$ . Legyen  $T_w$  magassága  $k + 1$ . Ekkor a fa minden  $i$ -edik szintjén legfeljebb  $2^i$  db nemterminális állhat ( $i = 1, \dots, k$ ), ha minden lépésben  $A \rightarrow BC$  szabályokat alkalmaztunk. Ebben az esetben, ha a  $k + 1$  edik szinten csak terminálisok állnak, akkor  $|w|$  hossza  $2^k = 2^{k+1-1}$ . Látható, hogy, ha valamelyik  $j$ -edik szinten ( $j < k$ ) egy  $A \rightarrow a$  szabályt alkalmazunk, akkor  $w$  hossza egyértelműen kisebb, mint  $2^k$ .  $\square$

## 6.5. Lineáris nyelvek

A lineáris nyelvek a környezetfüggetlen nyelvek egy részhalmazát képezik. Csak  $A \rightarrow \alpha B \beta$  és az  $A \rightarrow \alpha$  alakú szabályok generálják őket, ahol  $A, B \in V_N$ ,  $\alpha, \beta \in V_T^*$ , tehát a szabály jobb oldalán maximum egy nemterminális szerepelhet.

## 7. Iterációs lemmák

Iterációs lemmák (másnéven pumpáló lemmák) segítségével tudjuk eldönteni egy adott nyelvről, hogy egy adott nyelvosztályba tartozás feltételeit kielégíti-e, vagy sem. Ezeket a feltételeket szokás pumpálhatósági feltételeknek nevezni.

**7.1. Definíció.** Legyen  $V$  egy ábécé,  $L \subseteq V^*$  egy végtelen számosságú nyelv és  $x \in V^*$  egy  $V$  ábécé feletti szó, mely felírható  $x = uvw$  alakban. Azt mondjuk, hogy  $v$  pumpálható  $x$ -ben, akkor és csak akkor, ha minden  $i \geq 0$ -ra:

$$uv^i w \in L \Leftrightarrow x \in L.$$

**7.2. Definíció.** Legyen  $x \in V^*$  egy szó, ekkor  $d(x)$  jelöli  $x$  megkülönböztetett pozícióinak a számát,  $e(x)$  pedig  $x$  kizárt pozícióinak számát. Egy kizárt pozíció lehet megkülönböztetett is. Továbbá legyen  $s(x)$  az  $x$  beli kiválasztott, vagyis azon megkülönböztetett pozíciók száma, melyek nem kizárt pozíciók.

*Megjegyzés.* Triviális, hogy  $d(x) \leq s(x) + e(x)$ .

Az előbbi definíciók ismeretében a teljesség igénye nélkül vizsgálunk néhány iterációs lemmát.

### 7.1. Iterációs lemmák reguláris nyelvekre

Fentebb kimondtuk, hogy a véges automaták által felismert nyelvek osztálya reguláris. Legyen  $V$  egy ábécé és  $L \subseteq V^*$  egy végtelen számosságú reguláris nyelv. Tegyük fel, hogy  $L$ -t az  $A = (Q, V, \delta, q_0, F)$  determinisztikus véges automata ismeri fel, tehát  $L = L(A)$ . Ha veszünk egy  $w \in L$  szót, melyre igaz, hogy  $|w| > |Q|$ , akkor könnyen látható, hogy az  $A$  automata csak akkor ismeri fel  $w$ -t, ha működése során legalább egy állapotot többször is felvesz. Ebből az következik, hogy  $w$ -nek van egy bizonyos feltételeknek eleget tevő rész-szava, amely végtelen sokszor ismétlődhet és az ismétlődésekkel keletkezett szó is benne van az  $L$  nyelvben.

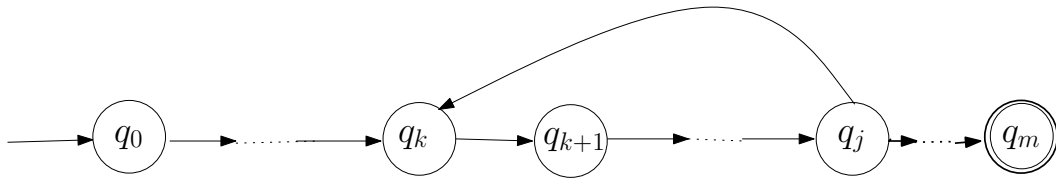
**7.1. Lemma** (Determinisztikus véges automatákra vonatkozó általános iterációs lemma). *Legyen  $A = (Q, V, \delta, q_0, F)$  egy determinisztikus véges automata. Ekkor létezik egy  $n > 0$  egész konstans (amelyre általában igaz, hogy  $n \leq |Q|$ ) úgy, hogy minden  $x \in L(A)$ , melyre teljesül  $|x| \geq n$ , felírható  $uvw$  alakban, ahol  $|v| \neq 0$ ,  $|uv| \leq n$ , továbbá  $uv^i w \in L(A)$  minden  $i \geq 0$ -ra.*

A következő lemma az előbbinek a reguláris nyelvekre megfogalmazott változata.

**7.2. Lemma** (Általános iterációs lemma reguláris nyelvekre [6]). *Legyen  $V$  egy ábécé. Minden  $L \subseteq V^*$  reguláris nyelvhez létezik egy  $n > 0$  egész konstans úgy, hogy minden  $x \in L$ , melyre teljesül  $|x| \geq n$ , felírható  $uvw$  alakban, ahol*

- $v \neq \lambda$ ,
- $|uv| \leq n$  és
- $uv^i w \in L$  minden  $i \geq 0$ -ra.

*Bizonyítás.* Legyen  $L$  a vizsgált nyelv. Tegyük fel, hogy  $L$  reguláris. Ekkor létezik egy  $A = (Q, V, \delta, q_0, F)$  determinisztikus véges automata, amelyre  $L(A) = L$ , továbbá legyen  $n > |Q|$  és  $x \in L$ ,  $|x| = m \geq n$ ,  $x = x_1 x_2 \dots x_m$ . Az  $A$  automata az  $x$  szót felismeri véges sok lépésben, közben a  $q_0, q_1, \dots, q_m \in Q$  belső állapotokat veszi fel. Mivel  $n > |Q|$ , ezért létezik  $k, j$  ( $0 \leq k \leq m$ ,  $0 \leq j \leq m$ ) számpár, melyre teljesül, hogy  $q_k = q_j$ , ( $k < j$ ). Továbbá  $k, j$ -t meg lehet választani úgy, hogy a  $q_{k+1}, q_{k+2}, \dots, q_j$  sorozat elemei páronként különbözőek legyenek (2. ábra).



2. ábra. Az automata működését reprezentáló gráf.

Legyen  $u = x_1 x_2 \dots x_k$ ,  $v = x_{k+1} x_{k+2} \dots x_j$  és  $w = x_{j+1} x_{j+2} \dots x_m$ . Ha  $k = 0$ , akkor  $u = \lambda$ , ha  $j = m$ , akkor  $w = \lambda$ . Mivel az előbb kikötöttük, hogy  $k < j$ , ezért  $v \neq \lambda$ , így az  $|uv| \leq n$  teljesül. Az automata felismeri a szót, tehát:

$$(q_0, x) \vdash^* (q_m, \lambda) \quad \text{másképp:}$$

$$(q_0, uvw) \vdash^* (q_k, vw) \vdash^* (q_k, w) \vdash^* (q_m, \lambda)$$

Mivel  $(q_k, vw) \vdash^* (q_k, w)$ , látható, hogy  $(q_k, v) \vdash^* (q_k, \lambda)$ , így igaz az is, hogy

$$(q_i, v^i) \vdash^* (q_k, \lambda) \text{ minden } i \geq 0 \text{-ra, melyből következik}$$

$$(q_0, uv^i w) \vdash^* (q_k, v^i w) \vdash^* (q_k, w) \vdash^* (q_m, \lambda) \text{ minden } i \geq 0$$

tehát  $uv^i w \in L$ . □

Az általános iterációs lemma a reguláris nyelvek osztályába való tartozásnak csak a szükséges feltételét vizsgálja, az elégséges feltételt nem. Nézzünk egy példát olyan nyelvre, melyre meg tudjuk mutatni, hogy nem tartozik a reguláris nyelvek osztályába. Legyen  $V = \{a, b\}$  és  $L = \{a^j b^j \mid j \in \mathbb{N}\} \subseteq V^*$  a vizsgált nyelv. Könnyen belátható, hogy  $L$  nem reguláris, mert bármely  $x \in L$  szó  $uvw$  felírásában a  $v$  szó megválasztása 3 féle módon történhet:

1.  $v$  csak  $a \in V$  szimbólumokat tartalmaz
2.  $v$  csak  $b \in V$  szimbólumokat tartalmaz
3.  $v = a^k b^l$  alakú, ahol  $k, l \in \mathbb{N}$ .

Az első esetben legyen  $k = |v| > 0$  ekkor  $v = a^k$ , amiből következik, hogy  $uv^i w = a^{j+k \cdot i} b^j$ , amely nincs benne az  $L$  nyelvben. A második eset az elsőhöz hasonló azzal a különbséggel, hogy  $uv^i w = a^j b^{j+k \cdot i}$ . A harmadik esetben pedig  $v = a^k b^l$ ,  $k, l > 0$ ,  $|v| = k + l$ , melyre  $uv^i w = a^{j-k} (a^k b^l)^i b^{j-l}$ , amely szót nem tartalmazza az  $L$  nyelv.

Jeffrey Jaffe a kongruencia reláció tulajdonságait felhasználva a reguláris nyelvekre vonatkozó lemma olyan változatát fogalmazta meg, mely a reguláris nyelvek osztályába való tartozásnak mind az elégséges, mind a szükséges feltételét vizsgálja ([8]). A lemma a következő.

**7.3. Lemma.** *Legyen  $L \subseteq V^*$  egy reguláris nyelv. Ekkor létezik egy  $n > 0$  egész konstans úgy, hogy minden  $z \in L$ , melyre  $|z| \geq n$ , felírható  $z = uvw$  alakban, melyre teljesülnek a következők:*

- $|v| \geq 1$
- $|uv| \leq n$
- minden  $i \geq 0$  -ra: minden  $x \in V^*$  -ra  $zx \in L \Leftrightarrow uv^i wx \in L$ .

□

A lemma bizonyítása a 7.2. lemma bizonyításából, valamint a jobb-kongruencia reláció természetéből egyértelműen belátható.

## 7.2. Iterációs lemmák környezetfüggetlen nyelvekre

**7.4. Lemma** (Lineáris nyelvekre vonatkozó iterációs lemma [7]). *Ha egy adott  $L$  nyelv lineáris, akkor létezik egy  $n$  egész szám úgy, hogy bármely olyan  $z \in L$  szó, amelyre  $|z| \geq n$ , felírható  $z = uvwxy$  alakban, ahol:*

$$vx \neq \lambda, \quad |uvxy| \leq n, \quad uv^i wx^i y \in L \quad \text{minden } i \geq 0 \text{ -ra.}$$

**Bizonyítás.** Legyen  $L$  egy lineáris nyelv, melyet a  $G = (V_N, V_T, S, H)$  grammatika generál. Ha vannak  $H$  -ban láncszabályok, akkor elimináljuk őket. Ekkor minden  $A \rightarrow uBv \in H$  -ra  $uv \neq \lambda$  teljesül. Legyen  $k = |V_N|$ , és  $d$  egy olyan pozitív egész szám, hogy minden  $A \rightarrow w \in H$  szabályra  $|w| \leq d$ . Ekkor  $n = k \cdot d$  és válasszunk egy  $z \in L$ ,  $|z| \geq n$  szót. Ekkor  $z$  -nek a levezetése a következő:

$$S \Rightarrow^+ u_1 A_1 y_1 \Rightarrow^+ \dots \Rightarrow^+ u_1 \dots u_m A_m v_m \dots v_1 \Rightarrow^+ u_1 \dots u_m p v_m \dots v_1 = z,$$

ahol  $A_1, \dots, A_m \in V_N$ ,  $u_1, v_1, \dots, u_m, v_m, p \in V_T^*$ . Jelöljük  $S$  -t  $A_0$  -al, ekkor  $A_i = A_j$  valamely  $i, j$  ( $0 \leq i < j \leq m-1$ ) párra az  $A_0, A_1, \dots, A_{m-1}$  sorozatban. Továbbá legyen  $j$  a legkisebb olyan index, amelyre az előbbi teljesül.

Ha  $i = 0$ , legyen  $u = y = \lambda$ ,  $v = u_1 \dots u_j$ ,  $w = u_{j+1} \dots u_m p v_m \dots v_{j+1}$ ,  $x = v_j \dots v_1$ .

Ha  $i > 0$ , akkor legyen  $u = u_1 \dots u_i$ ,  $v = u_{i+1} \dots u_j$ ,  $w = u_{j+1} \dots u_1 p v_1 \dots v_{j+1}$ ,  $x = v_j \dots v_{i+1}$ ,  $y = v_i \dots v_1$ .

Egyértelmű, hogy  $|uvxy| \leq j \cdot d \leq n$ , valamint  $vx \neq \lambda$  teljesül mindkét esetben. Végül  $S \Rightarrow^* u A_j y$ ,  $A_j \Rightarrow^* v A_j x$ ,  $A_j \Rightarrow^* w$ , melyekből következik, hogy  $uv^i w x^i y \in L$  minden  $i \geq 0$  -ra.  $\square$

Nézzünk a lemma alkalmazására egy példát. Legyen a vizsgált nyelv a  $\{a, b\}$  ábécé feletti  $L = \{w \mid |w|_a = |w|_b\}$ .

**7.5. Tétel.** A  $V = \{a, b\}$  ábécé feletti  $L = \{w \in V^* \mid |w|_a = |w|_b\}$  nyelv nem lineáris.

**Bizonyítás.** Tegyük fel, hogy  $L$  lineáris, akkor létezik egy  $n$  egész szám úgy, hogy létezik  $z \in L$ ,  $|z| \geq n$  szó, melyre teljesülnek a lemmában kikötött feltételek. Legyen ez a szó  $z = a^n b^{2n} a^n$ . Egyértelmű, hogy  $z \in L$ . Adjuk meg a  $z$  szó  $uvwxy$  felbontását. A következő feltételeknek teljesülnie kell  $|uvwx| \leq n$ ,  $vx \neq \lambda$ . Mivel  $|uvwx| \leq n$ , a  $z$  szó  $vx$  rész szava nem tartalmazhat egyszerre  $a$  és  $b$  szimbólumokat. A következő felbontási lehetőségek vannak:

Ha  $vx$  csak  $a$  szimbólumokat tartalmaz, ekkor csak  $i = 1$  -re teljesül, hogy  $uv^i w x^i y \in L$ .

Ugyanez az eset áll elő, ha  $vx$  csak  $b$  szimbólumokat tartalmaz.

Tehát az  $L$  nyelv nem teljesíti a lineáris iterációs lemma feltételeit, ezért az  $L$  nyelv nem lineáris.  $\square$

**7.6. Lemma** (Bar-Hillel Lemma [3]). Minden  $L$  környezetfüggetlen nyelvhez létezik egy ( $L$  -től függő)  $n$  pozitív egész szám úgy, hogy minden  $z \in L$ , melyre  $|z| > n$  teljesül, felírható  $uvwxy$  alakban, ahol

$$|vwx| \leq n, \quad vx \neq \lambda, \quad \text{és} \quad uv^i w x^i y \in L \quad \text{minden} \quad i \geq 0 \text{-ra.}$$

*Bizonyítás.* Legyen  $L$  egy környezetfüggetlen nyelv, melyet egy  $G = (V_N, V_T, S, H)$  láncszabály-mentes környezetfüggetlen nyelvtan generál. Legyen  $d$  egy olyan egész szám, melyre minden  $A \rightarrow w \in H$  szabályban  $|w| \leq d$  teljesül. Valamint  $k = |V_N|$ . Válasszuk  $n$ -t  $d^{k+1}$ -nek és legyen  $z \in L$  olyan szó, melyre  $|z| > n$ . Legyen  $T_z$  a  $z$  szó legrövidebb levezetési fája.  $T_z$  minden levéleme vagy  $\lambda$ , vagy terminális szimbólum, a belső elemek pedig nemterminális szimbólumok. Az  $n$  konstans megválasztásából következik, hogy  $T_z$ -ben van legalább egy olyan út, melynek hossza legalább  $k + 1$ . Jelölje  $t$  az egyik leghosszabb utat  $T_z$ -ben. Ekkor  $t$  legalább  $k + 2$  csomópontot köt össze, melyekből a legutolsó terminális, az összes többi nemterminális szimbólumot jelöl. Ebből következik, hogy legalább egy nemterminális kétszer szerepel az úton. Jelölje  $t'$  azt a legrövidebb  $t$ -beli utat, melyben történik ez az ismétlődés. Jelöljük  $t'$  gyökerét  $A$ -val. Felbontjuk a  $T_z$  fát három részfára:  $T_{uAy}, T_{vAx}, T_w$ , melyből megkapjuk  $u, v, w, x, y \in V_N^*$ -beli szavakat (3. ábra). Mutassuk meg, hogy erre a felbontásra van szükségünk.

Legyen  $v = x = \lambda$ . Ekkor  $S \Rightarrow^* z$  levezetés rövidebb, mint az eredeti (mely  $z$ -nek a legrövidebb levezetése), ami ellentmondáshoz vezet. Ebből következik, hogy  $vz \neq \lambda$ .

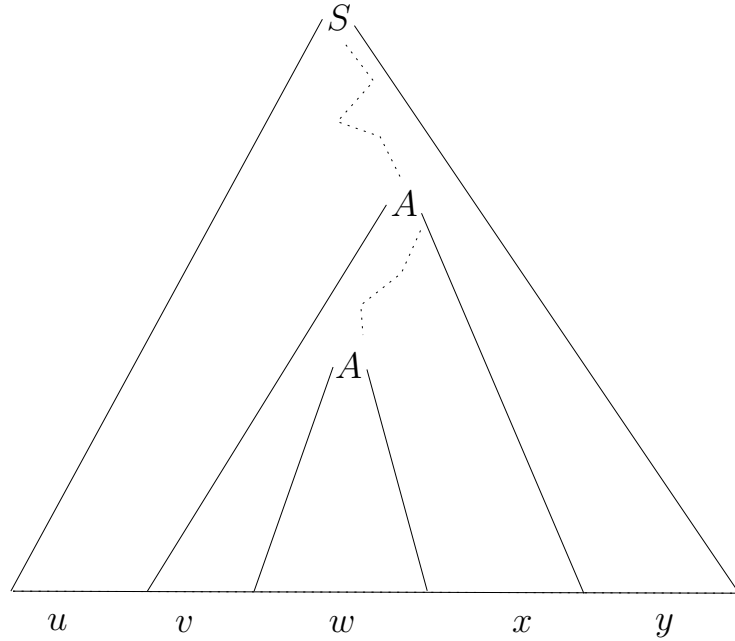
Mivel  $t$  az egyik leghosszabb út  $T_z$ -ben,  $T_{vAx}$ -ben egyetlen út sem lehet hosszabb, mint  $t'$ , továbbá  $t'$  hossza legfeljebb  $k + 1$ , melyekből következik, hogy  $|vwx| \leq d^{k+1} \leq n$ .

Végül  $i = 1$ -re  $uv^iwx^i y \in L$  egyértelműen teljesül. Ha  $i = 0$ , akkor a  $T_{vAx}$  részfa törlésével  $uwy \in L$  is fennáll. Valamint  $i \geq 2$  elérhető  $T_{vAx}$  részfa másolatainak  $A$  csomópontokhoz való hozzáfűzésével. Tehát  $uv^iwx^i y \in L$  teljesül minden  $i \geq 0$ -ra.  $\square$

Nézzük meg a 7.6 lemma alkalmazását. A legáltalánosabb példa erre az  $\{a, b, c\}$  ábécé feletti  $\{a^j b^j c^j \mid j \geq 1\}$  nyelv. Köztudott, hogy ez a nyelv nem környezetfüggetlen. Ezt bebizonyítjuk a Bar-Hillel lemma segítségével.

**7.7. Tétel.** Az  $L = \{a^j b^j c^j \mid j \geq 1\}$  nyelv nem környezetfüggetlen.

*Bizonyítás.* Tegyük fel, hogy  $L$  környezetfüggetlen, ekkor alkalmazható rá a Bar-Hillel lemma. Legyen  $n$  a 7.6. lemmában megfogalmazott egész konstans és  $z = a^n b^n c^n$ . Látható, hogy  $|z| = 3 \cdot n \geq n$ . Most írjuk fel az  $z$  szó  $uvwxy$  felbontását. Mivel  $|vwx| \leq n$ ,  $vx$  nem tartalmazhat  $a$ -kat és  $c$ -ket is, mivel az  $a$ -k és  $c$ -k között  $n$  db  $b$  áll. Így  $vx$ -et 5 módon választhatjuk meg. Ha  $vx$  csak  $a$ -kat tartalmaz, akkor  $uv^iwx^i y$   $i = 0$ -ra  $n$ -nél kevesebb  $a$ -t tartalmaz, a  $b$ -k és  $c$ -k száma pedig pontosan  $n$ . Ez a szó pedig nincs benne az  $L$  nyelvben. Ha  $vx$  csak  $c$ , vagy csak  $b$  szimbólumokat tartalmaz, akkor hasonlóan az előzőhöz,  $p$ -nél kevesebb  $c$ , vagy  $b$  szimbólum lesz a szóban. Tehát ezekben az esetekben nem teljesülnek a lemmában megfogalmazott feltételek, így  $L$  nem környezetfüggetlen. Választhatjuk  $vx$ -et úgy, hogy  $a$  és  $b$  szimbólumokat tartalmaz, ekkor  $i = 0$ -ra  $uv^iwx^i y$ -ban  $n$ -nél kevesebb lesz az  $a$ -k és  $b$ -k



3. ábra.  $z = uvwxy$  szó levezetési fája.

száma. Hasonló a helyzet, ha  $vx$   $b$  és  $c$  szimbólumokat tartalmaz. Ezzel bebizonyítottuk, hogy az  $L = \{a^j b^j c^j | j \geq 1\}$  nyelv nem környezetfüggetlen.  $\square$

Nézzünk még egy általános példát. Legyen a vizsgált nyelv  $L = \{a^{k^2} | k \geq 0\}$ .

**7.8. Tétel.** Az  $L = \{a^{k^2} | k \geq 0\}$  nyelv nem környezetfüggetlen.

*Bizonyítás.* Tegyük fel, hogy  $L$  környezetfüggetlen, ekkor a Bar-Hillel lemma feltételeinek elég kell tennie. Legyen  $z = a^{n^2}$ . Teljesül, hogy  $|z| \geq n$ . Az  $n^2$  utáni legkisebb teljes négyzet  $(n+1)^2 = n^2 + 2n + 1$ . De  $vx$  maximum csak  $n$  darab  $a$ -t tartalmazhat, így  $i = 2$ -re  $uv^2wx^2y =$  csak  $n^2 + 2n$  darab  $a$ -t tartalmaz, tehát nincs benne  $L$ -ben. Ebből következik, hogy az  $L$  nyelv nem környezetfüggetlen.  $\square$

Vannak olyan nem környezetfüggetlen nyelvek, melyek a Bar-Hillel lemma feltételeit kielégítik így szigorúbb megszorításokkal kell vizsgálni őket. A következő Ogden lemma a Bar-Hillel lemmánál szűkebb feltételeket szab meg.

**7.9. Lemma** (Ogden lemma [12]). Legyen  $G = (V_N, V_T, S, H)$  környezetfüggetlen nyelvtan,  $k = |V_N|$ , valamint  $l$  olyan egész szám, hogy minden  $A \rightarrow w$ -re  $|w| \leq l$ . Ekkor létezik egy  $n = l^{k+1}$  pozitív egész konstans úgy, hogy bármely  $z \in L(G)$ ,  $|z| \geq n$  szóra, ha  $d(z) \geq n$ , vagyis a  $z$  szó legalább  $n$  szimbólumát megkülönböztetettnek jelöljük, akkor  $z$  felírható  $z = uvwxy$  alakban, amire teljesülnek a következők:

1.  $d(vx) \geq 1$ , vagyis  $vx$  -ben legalább egy megkülönböztetett pozíció van
2.  $d(vwx) < n$ , vagyis  $vwx$  -ben maximum  $n$  db megkülönböztetett pozíció van
3. Létezik egy  $A \in V_N$  nemterminális, melyre

$$S \Rightarrow^* uAy, \quad A \Rightarrow^* vAx, \quad \text{továbbá} \quad A \Rightarrow^* w$$

tehát minden  $i \geq 0$  -ra  $uv^iwx^i y \in L(G)$

Vagy környezetfüggetlen nyelvekre átfogalmazva:

**7.10. Lemma** (Ogden lemma [12]). *Bármely  $L$ , környezetfüggetlen nyelvhez létezik egy  $n \geq 0$  pozitív egész konstans úgy, hogy bármely  $z \in L$ ,  $|z| \geq n$  szóra, ha  $d(z) \geq n$ , vagyis a  $z$  szó legalább  $n$  szimbólumát megkülönböztetettnek jelöljük, akkor  $z$  felírható  $z = uvwxy$  alakban, amire teljesülnek a következők:*

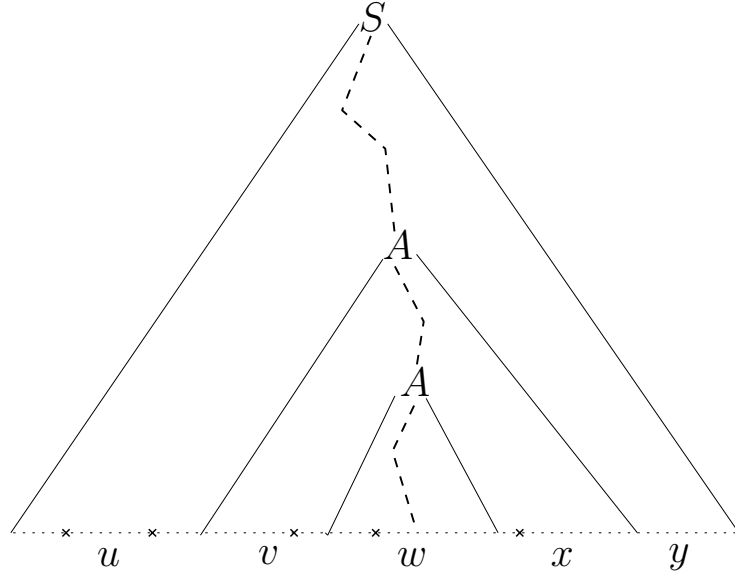
1.  $d(vx) \geq 1$ , vagyis  $vx$  -ben legalább egy megkülönböztetett pozíció van
2.  $d(vwx) \leq n$ , vagyis  $vwx$  -ben maximum  $n$  db megkülönböztetett pozíció van
3. minden  $i \geq 0$  -ra  $uv^iwx^i y \in L$

*Bizonyítás.* Legyen  $G = (V_N, V_T, S, H)$  egy környezetfüggetlen nyelvtan és  $z \in L$ , melyre teljesül  $|z| \geq n = l^{k+1}$ . Legyen  $T_z$  a  $z$  szó levezetési fája. Minden  $c$  csomópont  $d(c)$  jelenti  $c$  leszármazotaiban lévő csomópontok darabszámát. A következő eljárással előállítunk egy  $T_z$  -beli  $t$  utat, melyben minden  $c$  csomópont  $d(c) > 0$ .

Adjuk  $T_z$  gyökerét  $t$  -hez. Legyen  $c$  az utolsó csomópont, amit  $t$  -hez adtunk, ekkor vizsgáljuk  $c$  gyermekeit. Ha  $c$  -nek nincs több gyermeke, akkor megállunk. Ha  $c$  -nek egyetlen gyermeke van, amit jelöljünk  $b$  -vel és  $d(b) > 0$ , adjuk  $b$  -t  $t$  -hez és hajtsuk végre az eljárást  $b$  -re. Egyébként, ha  $c$  -nek kettő, vagy több gyermekének van megkülönböztetett pozíciókat tartalmazó leszármazottja, akkor  $c$  -t elágazási pontnak nevezzük. Válasszuk  $c$  -nek olyan  $b$  gyermekét, melyre  $d(b)$  maximális. Adjuk hozzá ezt a  $b$  -t  $t$  -hez és folytassuk az eljárást.

Jelöljük  $t$  elágazási pontjait felülről-lefelé haladva  $c_0, c_1, \dots, c_j$  -vel. Könnyen látható, hogy  $d(c_{i+1}) \geq \frac{1}{l}d(c_i)$  minden  $0 \leq i < j$  -re. Egyértelmű, hogy  $d(c_0) \geq n$ , mivel  $d(z) \geq n$ . Kimondtuk, hogy  $n = l^{k+1}$ . Ekkor  $d(c_1) \geq \frac{1}{l}d(c_0) \geq l^{k+1-1}$ , vagyis  $d(c_i) \geq \frac{1}{l}d(c_{i+1}) \geq l^{k+1-i}$  minden  $0 \leq i < j$ , valamint  $d(c_j) \geq l^{k+1-j}$ . De  $l$  megválasztásából következik, hogy  $d(c_j) \leq l$ , mivel  $c_j$  -nek nincsenek olyan rákövetkezői, melyek elágazási pontok lennének. Ebből következik, hogy  $l \geq l^{k+1-j}$ , továbbá  $1 \geq k+1-j$ , így  $j \geq k$ .

Ezekből következik, hogy  $t$  legalább  $k+1$  elágazási pontot tartalmaz és mivel minden elágazási pont nemterminálist jelöl, legalább egy nemterminális kétszer fordul elő  $t$  -ben. Jelöljük



4. ábra.  $T_z$  fa. Szaggatott vonal jelöli a  $t$  utat, \* pedig a megkülönböztetett pozíciókat (más megkülönböztetett pozíciók is előfordulhatnak).

ez a nemterminálist  $A$  -val. Ekkor léteznek  $S \Rightarrow^* uAy$ ,  $A \Rightarrow^* vAx$  és  $A \Rightarrow^* w$  levezetések (4. ábra).

Látható, hogy  $vwx$  a  $T_z$  egy részfájának levélelemeiből áll össze, melynek gyökerét jelöljük  $c_i$  -vel. Továbbá  $c_i$  a felső  $A$  nemterminálist jelöli és ezt az elágazási pontot maximum  $k$  rákövetkező elágazási pont követheti  $t$  -ben. Tudjuk, hogy  $d(c_j) \leq l$ ,  $d(c_{j-m}) \leq l^{1+m}$  minden  $j \leq m \leq i$  -re. Tehát  $d(c_i) \leq l^{k+1}$ . Ezzel láthatjuk, hogy  $d(vwx) \leq l^{k+1} = n$ . Mivel  $c_i$  a felső  $A$  elágazási pontot jelölte, tudjuk, hogy  $c_i$  -nek van legalább két gyermeke, melyek közül az egyiknek a leszármazottja  $w$ , a másiknak a leszármazottja legalább egy  $vx$  -beli szimbólum.

Ha  $i = 0$ , akkor  $S \Rightarrow^* uAy \Rightarrow^* uwy$  levezetés fennáll és  $uwy \in L(G)$  teljesül.  $i = 1$  -re pedig egyértelmű, hogy  $z = uv^iwx^i y \in L(G)$ . Ha  $i \geq 2$ , akkor  $A \Rightarrow^* vAx$  levezetést többször alkalmazva kapjuk  $uv^iwx^i y \in L(G)$  szót.  $\square$

Az Ogden lemma alkalmazását nézzük meg egy olyan nyelvre, melyre a Bar-Hillel lemma feltételei nem nyújtanak elegendő megszorítást. Legyen a vizsgált nyelv

$$L = \{a^i b^j c^k \mid i, j, k \geq 0 \text{ és vagy } i = j, \text{ vagy } j = k, \text{ de nem } i = j = k\}.$$

Így, ha  $z$  -t valamilyen  $a^n b^n c^l$  -nek választjuk, ahol  $l \neq n$ ,  $vx$  -et választhatjuk úgy, hogy csak  $c$  -ket tartalmazzon és bármely  $m \geq 0$  -ra teljesül, hogy  $uv^m wx^m y \in L$ . Ugyanez a lehetőség fennáll, ha  $z = a^l b^n c^n$ , ekkor  $vx$  csak  $a$  -kat tartalmazhat. Így a Bar-Hillel lemma segítségével nem tudjuk bebizonyítani, hogy  $L$  nem környezetfüggetlen. Nézzük most, hogy az Ogden lemma alkalmazásával milyen eredmények érhetők el.

**7.11. Tétel.** Az  $L = \{a^i b^j c^k \mid i, j, k \geq 0 \text{ és vagy } i = j, \text{ vagy } j = k, \text{ de nem } i = j = k\}$  nyelv környezetfüggetlen.

*Bizonyítás.* Tegyük fel, hogy  $L$  környezetfüggetlen. Ekkor az Ogden lemmát alkalmazva válasszuk  $z$ -t  $a^n b^n c^{(n+n!)}$ -nak. Jelöljük az  $a$ -kat megkülönböztetettnek. A  $z$  szó  $uvwxy$  felbontásában ekkor  $vx$  legalább egy megkülönböztetett pozíciót tartalmaz, mely a következő lehet:

(I)  $v = \lambda$  és  $x$  tartalmaz egy  $a$ -t

(II)  $v$  tartalmaz egy  $a$ -t

Az I. esetben  $x$  nem tartalmazhat egyetlen  $b$ -t és egyetlen  $c$ -t sem, ellenkező esetben nem teljesül, hogy  $uv^i wx^i y \in L$ . Legyen  $x = a^l$  valamilyen  $l \geq 0$ -ra. Ekkor  $uv^i wx^i y = a^{(n+(i-1) \cdot l)} b^n c^{(n+n!)} \notin L$ .

A II. esetben  $v$  nem tartalmazhat  $b$ , vagy  $c$  szimbólumokat, ellenkező esetben  $uv^i wx^i y \notin L$ . Legyen  $v = a^l$  valamilyen  $1 \leq l \leq n$ -re. Továbbá  $x$  is csak egy fajta szimbólumot tartalmazhat, mert ellenkező esetben  $uv^i wx^i y \notin L$ . Tehát  $x = a^m$ , vagy  $x = b^m$ , vagy  $x = c^m$  valamilyen  $m$ -re. Ha  $x = a^m$ , vagy  $x = c^m$ , akkor  $uv^i wx^i y \notin L$ . Ha  $x = b^m$ , akkor vagy  $m \neq l$ , vagy  $m = l$ . Ha  $m = l$ , akkor  $i = 0$ -ra  $|uvy|_a \neq |uvy|_b$ , tehát  $uvy \notin L$ . Ha  $m = l$ , akkor  $i = \frac{n!}{l} + 1$ -re  $uv^{(\frac{n!}{l}+1)} wx^{(\frac{n!}{l}+1)} y = a^{(n+n!)} b^{(n+n!)} c^{(n+n!)} \notin L$ .

Ezzel bebizonyítottuk, hogy  $L$  nem környezetfüggetlen.  $\square$

A Bar-Hillel lemma feltételeit erősíthetjük Ogden feltételei alapján. Az így kapott lemma a következő.

**7.12. Lemma** (Erős Bar-Hillel lemma). Minden  $G = (V_N, V_T, S, H)$  környezetfüggetlen nyelvtanhoz létezik egy  $n$  pozitív egész szám úgy, hogy bármely  $z \in L(G)$ ,  $|z| > n$  szóban tetszőlegesen megkülönböztetettnek jelölünk legalább  $n$  pozíciót, akkor létezik  $A \in V_N$ , szimbólum és  $u, v, w, x, y \in V_T^*$  szavak úgy, hogy a következők teljesülnek:

- $S \Rightarrow^* uAy$ ,  $A \Rightarrow^+ vAx$ ,  $A \Rightarrow^+ w$ ,  $z = uvwxy$ ,  $|w| > 0$ ,
- vagy  $|u| > 0$  és  $|v| > 0$ , vagy  $|x| > 0$  és  $|y| > 0$ ,
- $d(vwx) \leq n$ ,
- $uv^i wx^i y \in L(G)$  minden  $i \geq 0$ -ra.

$\square$

Valamint a reguláris nyelvekre vonatkozó általános iterációs lemma feltételeit is át lehet alakítani Ogden lemmájának segítségével.

**7.13. Lemma.** Minden  $L$  reguláris nyelvhez létezik egy  $n > 0$  pozitív egész szám úgy, hogy bármely  $z \in L$ ,  $|z| \geq n$  szó melyben legalább  $n$  pozíciót megkülönböztettként jelölünk, felírható  $z = uvw$  alakban, melyre teljesülnek a következők:

- $d(uv) \leq n$
- $d(v) \geq 1$
- $uv^i w \in L$  minden  $i \geq 0$ -ra.

□

Az utóbbi két tétel bizonyítása az Ogden lemma után triviális, ezért nem szükséges vele foglalkozni.

A következő lemma segítségével pedig megmutathatjuk, hogy egy adott környezetfüggetlen nyelv nem lineáris.

**7.14. Lemma** (Iterációs lemma nem-lineáris környezetfüggetlen nyelvekre [7]). Legyen  $L$  egy nem-lineáris környezetfüggetlen nyelvtan, ekkor létezik (végtelen sok)  $p \in L$  szó, mely felírható  $p = rstuvwxyz$  alakban és teljesülnek a következők:

$$|su| \neq 0, \quad |wy| \neq 0, \quad rs^i t u^i v w^j x y^j z \in L \quad \text{minden } i, j \geq 0 \text{-ra.}$$

*Bizonyítás.* Legyen  $L$  egy környezetfüggetlen nyelv, melyet a  $G = (V_N, V_T, S, H)$  grammatika generál. Tegyük fel továbbá azt is, hogy  $G$  nem lineáris nyelvtan és  $H$  nem tartalmaz láncszabályokat. Létezik legalább egy olyan  $H$ -beli szabálya, melynek jobb oldalán legalább két nemterminális áll. Ekkor

$$S \Rightarrow^* x_1 A x_2 B x_3 \Rightarrow^* x_1 a A b x_2 c B d x_3 \Rightarrow^* o$$

alakú levezetés lehetséges, ahol  $a, b, c, d, x_i \in (V_N \cup V_T)^*$  minden  $i = 1, 2, 3$ -ra és  $A, B \in V_N$ . Továbbá  $L(A)$  és  $L(B)$  (tehát  $A$  és  $B$  nemterminálisokból  $H$ -beli szabályokkal levezethető szavak nyelve) végtelen számosságú, mert teljesül

$$A \Rightarrow^* aAb, \quad \text{és} \quad B \Rightarrow^* cBd, \quad \text{valamint} \quad A \Rightarrow^* g, \quad B \Rightarrow^* h \quad \text{ahol} \quad g, h \in V_T^*. \quad (1)$$

Most legyen  $p = o$  és az  $r, s, t, u, v, w, x, y, z$  szavakat a következőképpen válasszuk:

$$r = x_1, \quad s = a, \quad t = g, \quad u = b, \quad v = x_2, \quad w = c, \quad x = h, \quad y = d, \quad z = x_3.$$

Látható, hogy

$$rs^i t u^i v w^j x y^j z = x_1 a^i g b^i x_2 c^j h d^j x_3 \in L \quad \text{minden } i, j \geq 0 \text{-ra} \quad (1) \text{ miatt.}$$

Továbbá, mivel  $H$  nem tartalmaz láncszabályokat, minden  $A \rightarrow xBy$  szabályban  $|xy| \geq 1$ , tehát  $|su| \geq 1$  és  $|wy| \geq 1$  is teljesül. □

**7.15. Tétel.** Legyen  $G = (V_N, V_T, S, H)$  egy környezetfüggetlen nyelvtan,  $k = |V_N|$ ,  $l$  pedig olyan egész szám, hogy minden  $A \rightarrow w \in H$  -ra  $|w| \leq l$ . Ha egy  $z \in L(G)$  szó  $T_z$  levezetési fája legfeljebb  $m$  darab elágazási pontot tartalmaz, akkor  $d(T_z) \leq l^m$ .

*Bizonyítás.* Tegyük fel, hogy  $m = 0$ . Ekkor egyik belső csomópont sem elágazási pont, tehát  $d(T_z) \leq l^0 = 1$ . Ha  $d(T_z) = 2$  lenne, akkor a legalsó közös ő elágazási pont lenne.

Tegyük fel, hogy  $m-1$  -re igaz a tétel. Bizonyítsuk be, hogy  $m$  -re is igaz. Legyen  $T_z$  -ben  $m$  darab elágazási pont. Válasszunk ki az elágazási pontok közül egyet, melynek egyetlen őse sem elágazási pont és jelöljük  $q$  -val. Ekkor  $T_z$  minden megkülönböztetett levéleleme leszarmazottja  $q$  -nak. De  $q$  -nak legfeljebb  $l$  gyermeke lehet, melyek közül mindegyik valamilyen részfának a gyökere. Ezekben a részfákban az utak legfeljebb  $m-1$  elágazási pontot tartalmazhatnak. A feltevés szerint bármely ilyen részfa legfeljebb  $l^{m-1}$  elágazási pontot tartalmazhat. Ezekből következik, hogy  $d(T_z) \leq l^m$ .  $\square$

A következő lemma Ogden lemmájának egy bővített változata.

**7.16. Lemma** (Bader-Moura lemma [2]). Minden  $L$  környezetfüggetlen nyelvre létezik egy  $n$  egész szám úgy, hogy bármely  $z \in L$  szó, melyre tetszőlegesen kiválasztott megkülönböztetett pozíciók esetén teljesül  $d(z) > n^{e(z)+1}$  akkor  $z$  felírható  $uvwxy$  alakban, melyre teljesülnek a következők:

- $d(vx) \geq 1$  és  $e(vx) = 0$
- $d(vwx) \leq n^{e(vwx)+1}$
- $uv^iwx^iy \in L$  minden  $i \geq 0$  -ra.

*Bizonyítás.* Legyen  $L$  egy környezetfüggetlen nyelv, melyet egy  $G = (V_N, V_T, S, H)$  nyelvtan generál. Legyen  $k = |V_N|$  és  $l$  egy olyan pozitív egész szám, hogy minden  $A \rightarrow w$  -re  $l \geq |w|$ . Válasszuk  $n$  -t  $l^{k+1}$  -nek és legyen  $z \in L$ ,  $|z| \geq n$ . Legyen  $T_z$  a  $z$  szó levezetési fája és  $t$  a leghosszabb  $T_z$  -beli út. Mivel  $d(z) \geq l^{(k+1) \cdot (e(z)+1)}$ , a 7.15. tétel értelmében  $t$  legalább  $(k+1) \cdot (e(z)+1)$  elágazási pontot tartalmaz. Felosszuk  $t$  legalsó részét  $e(z)+1$  darab al-útra, melyek közül mindegyik  $k+1$  elágazási pontot tartalmaz. Minden ilyen al-útban lennie kell  $c_1, c_2$  pontoknak, melyek azonos nemterminálist jelölnek. Jelöljük ezt a nemterminálist  $A$  -val. Ekkor létezik  $v', x' \in V_T^*$  úgy, hogy  $A \Rightarrow^* v'Ax'$ . Továbbá létezik  $w' \in V_T^*$  úgy, hogy  $A \Rightarrow^* w'$ . A  $v', x'$  szavakat pumpálható rész-szavaknak nevezzük.

A  $c_1, c_2$  csomópontok azonos úton helyezkednek el és mindkettőnek két-két gyermekének vannak megkülönböztetett leszarmazottai. A felsőbbiknek legalább egy megkülönböztetett leszarmazottja van  $v'x'$  -ben, ebből következik, hogy  $v'x'$  tartalmaz legalább egy megkülönböztetett pozíciót.

A  $t$  termináló levéleleméből indulva lépkedjünk visszafelé az al-utakon amíg nem jutunk olyanhoz, melynek a pumpálható rész-szavai nem tartalmazznak kizárt pozíciókat. Tudjuk, hogy léteznek ilyen rész-szavak, mert  $e(z) + 1$  al-útban  $e(z) + 1$   $v', x'$  pár létezik, a kizárt pozíciók száma pedig csak  $e(z)$ . Az a pár, amit ezzel a módszerrel találunk, lesz  $v, x$ . Ezzel bebizonyítottuk, hogy  $d(vx) \geq 1$  és  $e(vx) = 0$ . Valamint az előbbi lemmákhoz hasonlóan, látható, hogy  $uv^iwx^iy \in L$  minden  $i \geq 0$ -ra.

Tegyük fel, hogy a  $g + 1$ -edik al-úton találtuk meg a  $v, x$  párt. Minden lentebbi al-úton  $e(v'x') \geq 1$ . Legyen  $s$  a  $g + 1$  alatti al-utakban lévő kizárt pozíciók száma. Látható, hogy  $s \geq g$ . A  $vwx$  levélelemekben végződő fa gyökere a  $g + 1$ -edik al-útban lévő egyik csomópont. A  $t$  út megválasztása miatt ebből a csomópontból induló utakban lévő elágazási pontok száma legfeljebb  $(k + 1) \cdot (g + 1)$ . Ezekből következik, hogy  $d(vwx) = r$ , ahol  $r \leq 2^{(k+1) \cdot (g+1)} \leq n^{g+1} \leq n^{s+1}$ , tehát  $d(vwx) \leq n^{e(vwx)+1}$ .  $\square$

Nézzük meg, hogy Bader és Moura feltételei hogyan alkalmazhatók egy olyan nyelvre, melyre az Ogden lemma nem nyújt elegendő megszorítást. Legyen a vizsgált nyelv

$$L = \{b^*\} \cup \{aa^+b^*\} \cup \{ab^p \mid p \text{ prím}\}.$$

Ha Ogden lemmáját alkalmazzuk,  $vx$ -et nem tudnánk úgy megválasztani, hogy az  $a$  szimbólumot ne tartalmazz. Tehát nem sikerülne bebizonyítani, hogy a nyelv nem környezetfüggetlen. Bader és Moura megszorításaival, ha olyan  $z$  szót választunk, mely  $ab^p$  alakú, ahol  $p$  prím és az  $a$  szimbólumot kizárt pozíciónak tekintjük, akkor  $vx$  csak  $b$ -ket tartalmaz és  $|uv^iwx^iy|_b$  nem biztos, hogy prím lesz, továbbá a szó egyetlen  $a$ -val kezdődik, tehát az  $aa^+b^*$  kifejezés által meghatározott nyelvben sincs benne. Ezzel bebizonyítottuk, hogy a fentebbi  $L$  nyelv nem környezetfüggetlen.

Egy, a lineáris nyelvekre vonatkozó iterációs lemmát is származtathatunk a Bader-Moura lemmából.

**7.17. Lemma** (Bader-Moura lemma általánosítása lineáris nyelvekre [4]). *Legyen  $L$  lineáris környezetfüggetlen nyelv, ekkor létezik egy  $n$  egész konstans úgy, hogy bármely  $z \in L$ , melyre tetszőlegesen megválasztott kizárt pozíciók esetén teljesül, hogy  $d(z) \geq n \cdot (e(z) + 1)$ , felírható  $z = uvwxy$  alakban, melyre teljesülnek a következők:*

- $d(vx) \geq 1$  és  $e(vx) = 0$
- $d(uvxy) \leq n \cdot (e(uvxy) + 1)$
- minden  $i \geq 0$ -ra  $uv^iwx^iy \in L$

$\square$

**7.18. Lemma** (Reguláris nyelvekre vonatkozó erős iterációs lemma (Bader-Moura lemma általánosítása) [5]). *Minden  $L$  reguláris nyelvhez létezik egy  $L$  -től függő  $n \geq 2$  egész konstans úgy, hogy bármely  $z \in L$ , melyre tetszőlegesen kiválasztott megkülönböztetett pozíciók esetén teljesül, hogy  $d(z) > n \cdot \max(e(z), 1)$ , felírható  $z = uvw$  alakban, melyre teljesülnek a következők:*

- $e(v) = 0, d(u) > 0, d(w) > 0$
- $d(uv) \leq n \cdot (e(u) + 1)$
- $uv^i w \in L$  minden  $i \geq 0$  -ra.

*Bizonyítás.* Legyen  $G = (V_N, V_T, S, H)$  egy reguláris grammatika, amelyre  $L = L(G)$  és minden  $H$  -beli szabály  $A \rightarrow Ba$ , vagy  $A \rightarrow a$  alakú. Ekkor minden levezetési fa lépcsőzetes alakú, ahol minden csomópont jobb oldali gyermeke terminális szimbólum, bal oldali gyermeke pedig nemterminális, kivéve az utolsó előtti csomópontot, mivel annak csak egy terminális leszármazottja van. Bármely  $z \in L$  -re  $T_z$  fa mélysége pontosan  $|z|$ .

Azok a csomópontok lesznek elágazási pontok, melyeknek a jobb oldali gyermekük kizárt pozícióban van.

Kiválasztott út lesz a levezetésnek olyan szakasza, melynek vagy a legfelső csomópontja a gyökér és a legalsó egy elágazási pont, vagy mind a felső, mind az alsó csomópont elágazási pont, vagy a felső csomópont gyökértől különböző elágazási pont és az alsó csomópont levél-elem és egyetlen köztes csomópont sem elágazási pont. Látható, hogy így  $d(z) \leq e(z) + 1$ .

Legyen  $k = |V_N|$  és  $n = 2 \cdot k + 5$ .

Ha  $e(z) = 0$ , akkor  $n$  -nél több megkülönböztetett pozíció van  $z$  -ben. Legyen  $T_z$  a  $z$  szó levezetése. Léteznek  $t_1, t_2, t_3$  utak  $T_z$  -ben úgy, hogy  $d(t_1) \geq 1, d(t_2) = k + 1, d(t_3) = 1$ . Ebből következik, hogy  $t_2$  -ben legalább egy nemterminális legalább két csomópontban fordul elő. Legyen ez a nemterminális  $A$ . Ekkor  $A \Rightarrow^* Av$  és  $A \Rightarrow^* u$  levezetések léteznek  $G$  -ben,  $d(v) > 0, z = uvw$  és minden  $i \geq 0$  -ra  $uv^i w \in L$ . Mivel  $d(t_3) = 1, d(u) > 0$ . Továbbá  $d(uv) \leq k + 2 \leq n$ .

Ha  $e(z) > 0$ , akkor nem teljesülhet, hogy minden megkülönböztetett útban legfeljebb  $k + 2$  kiválasztott pozíció legyen, mivel, ha teljesülne, akkor:

$$d(z) \leq (e(z) + 1) \cdot (k + 2) + e(z) \leq (2 \cdot k + 5) \cdot e(z)$$

Tehát létezik egy megkülönböztetett pozíciókat tartalmazó út (jelöljük  $t$  -vel), amelynek levél-elemeiben legfeljebb  $k + 1$  kiválasztott pozíció van. Ezeknek legalább  $k + 2$  őse van  $t$  -ben. Osszuk fel  $t$  -t  $t_1, t_2, t_3$  al-utakra, melyekre  $s(t_1) \geq 1, s(t_2) = k + 1, s(t_3) = 1$ . Ekkor  $t_2$  -ben egy  $A$  nemterminális kétszer fordul elő és  $A \Rightarrow^* Av, A \Rightarrow^* u$ , továbbá  $s(v) > 0, z = uvw$ ,

valamint minden  $i \geq 0$  -ra  $uv^i w \in L$ . Mivel  $s(t_3) = 1$ ,  $s(u) > 0$  és triviális, hogy  $e(v) = 0$ . Ha  $t$ -t úgy választjuk meg, hogy a legalsó olyan út legyen, melyre teljesülnek az előzők, akkor

$$d(uv) \leq (e(u) + 1) \cdot (k + 2) + e(u) \leq n \cdot (e(u) + 1).$$

□

**7.19. Tétel.** Legyenek  $L_1$  és  $L_2$  azonos típusú nyelvek (reguláris, vagy környezetfüggetlen). Ha valamelyik iterációs lemma feltételeit kielégíti  $L_1$ , valamint  $L_2$  is, akkor  $L_1 \cup L_2$  is kielégíti a lemma feltételeit.

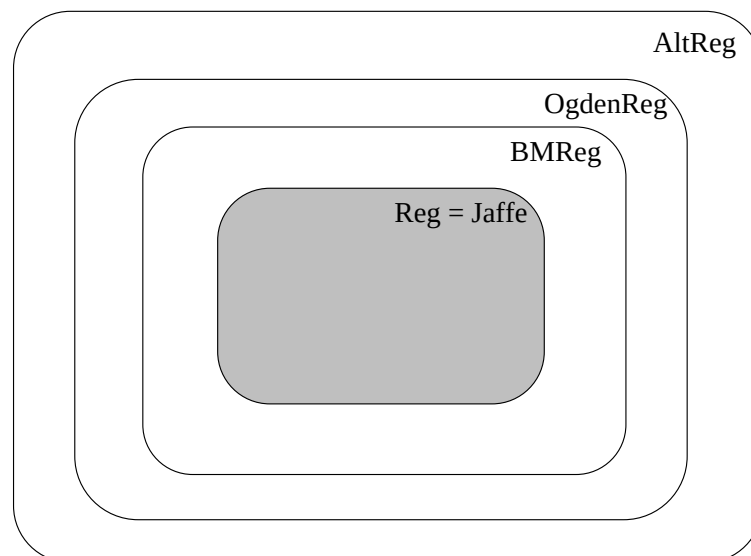
*Bizonyítás.* Könnyen belátható. Legyen  $n_1, n_2$  rendre az  $L_1, L_2$  nyelvhez rendelt konstansok. Ekkor az  $L_1 \cup L_2$  nyelvhez rendelt konstans,  $n$ , legyen  $\max(n_1, n_2)$ . □

## 8. Összefoglalás

Megismertük tehát a reguláris és környezetfüggetlen nyelvek egyes jellemzőit, az őket elfogadó automatókat és a nyelvekre vonatkozó legelterjedtebb iterációs lemmákat.

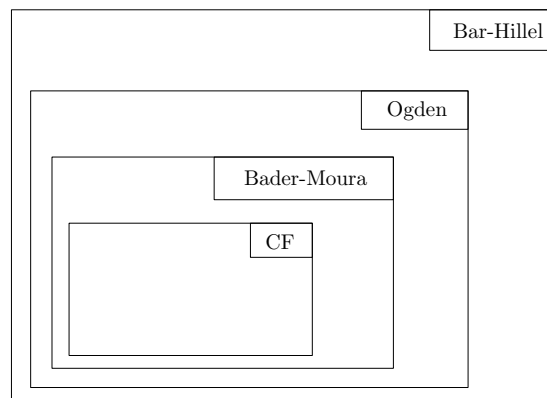
A reguláris nyelvekre vonatkozó általános iterációs lemma nem nyújt elég megszorítást ahhoz, hogy bármely nyelvről be lehessen bizonyítani, hogy nem reguláris. Tehát a 7.2. lemma által meghatározott nyelvek valódi részhalmaza a reguláris nyelvek osztálya. Jeffrey Jaffe a Nerode-féle kongruencia segítségével kidolgozott elegáns lemmája pontosan a reguláris nyelveket határozza meg. A következő ábrában láthatjuk az egyes reguláris iterációs lemmák által lefedett nyelvek hierarchiáját, ahol

- *Reg* a reguláris nyelvek osztályát jelöli,
- *Jaffe* Jeffrey Jaffe lemmája által meghatározott nyelvek halmaza,
- *BMReg* Bader-Moura lemma reguláris változata által meghatározott nyelvek,
- *OgdenReg* Ogden lemmájának reguláris nyelvekre általánosított lemmája által meghatározott nyelvek,
- *AltReg* pedig a reguláris nyelvekre vonatkozó általános iterációs lemma által meghatározott nyelvek halmaza.



A környezetfüggetlen nyelvek esetében csak jóval bonyolultabb módszerrel lehet elégséges és szükséges feltételeket vizsgáló lemmát megfogalmazni. Ezért erre a problémára a dolgozatban nem tértem ki. A Bar-Hillel lemma, az általános reguláris iterációs lemmához hasonlóan,

egyszerű, könnyen érthető, de a környezetfüggetlen nyelveknél jóval nagyobb halmazt foglal magába. Ogden lemmájában a megkülönböztetett pozíciók segítségével a pumpálható rész-szóra adhatunk megszorításokat. Így az általa lefedett halmaz kisebb a Bar-Hillel lemma által lefedett halmaznál, de a környezetfüggetlen nyelvek ennek a halmaznak is csak valódi részhalmazát képezik. Bader és Moura lemmája a kizárt pozíciók segítségével a pumpálható rész-szóra több megszorítást ad, ezzel tovább szűkül a lefedett halmaz. A következő ábra szemlélteti az egyes környezetfüggetlen iterációs lemmák által meghatározott nyelvek hierarchiáját, ahol az egyes lemmák által meghatározott halmazok a lemma nevével vannak jelölve, a környezetfüggetlen nyelvek halmaza pedig  $CF$  -el.



## **9. Köszönetnyilvánítás**

Köszönettel tartozom témavezetőmnek, Dr. Dömösi Pálnak, aki érdekes előadásaival, sok anyaggal és tanácsokkal segítette munkámat, valamint Horváth Gézának, Dr. Nagy Benedeknek és Giora Slutzki professzornak, akik szintén anyagokkal segítették munkámat.

Külön köszönet illeti édesanyámat, aki mindig támogatott tanulmányaimban.

## Hivatkozások

- [1] Alfred V. Aho and Jeffrey D. Ullman. *The theory of parsing, translation, and compiling*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1972.
- [2] Christopher Bader and Arnaldo Moura. A generalization of ogden’s lemma. *J. ACM*, 29(2):404–407, 1982.
- [3] Yehoshua Bar-Hillel, M. Perles, and E. Shamir. On formal properties of simple phrase structure grammars. *Zeitschrift für Phonologie, Sprachwissenschaft und Kommunikationsforschung*, 14:113–124, 1961.
- [4] R. Boonyavatana and G. Slutzki. The interchange or pump (di)lemmas for context-free languages. *Theor. Comput. Sci.*, 56(3):321–338, 1988.
- [5] Pál Dömösi and Manfred Kudlek. Strong iteration lemmata for regular, linear, context-free, and linear indexed languages. In *FCT ’99: Proceedings of the 12th International Symposium on Fundamentals of Computation Theory*, pages 226–233, London, UK, 1999. Springer-Verlag.
- [6] John E. Hopcroft and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading, Massachusetts, 1979.
- [7] Géza Horváth. New pumping lemma for not k-linear context-free languages. 2009.
- [8] Jeffrey Jaffe. A necessary and sufficient pumping lemma for regular languages. *SIGACT News*, 10(2):48–49, 1978.
- [9] William F. Ogden. A helpful result for proving inherent ambiguity. *Mathematical Systems Theory*, 2(3):191–194, 1968.
- [10] Arto Salomaa. *Computation and automata*. Cambridge University Press, New York, NY, USA, 1985.
- [11] Arto Salomaa. *Formal languages*. Academic Press Professional, Inc., San Diego, CA, USA, 1987.
- [12] Jeffrey Shallit. *A Second Course in Formal Languages and Automata Theory*. Cambridge University Press, New York, NY, USA, 2008.
- [13] D. S. Wise. A strong pumping lemma for context-free languages. *Theor. Comput. Sci.*, (3):359–369, 1976.