4. Reguláris nyelvtanok, CF nyelvtanok és ε-szabályok

1. Igazolja, hogy minden nem reguláris $L_1 \subseteq \{\mathtt{a}, \mathtt{b}\}^*$ nyelvhez van olyan nem reguláris $L_2 \subseteq \{\mathtt{a}, \mathtt{b}\}^*$ nyelv, hogy az $L_1 \cup L_2$ nyelv reguláris!

   Megoldás: Ha L_1 egy tetszőleges nem reguláris nyelv, akkor a komplementere sem reguláris, legyen ez az L_2. Így $L_1 \cup L_2 = \{\mathtt{a}, \mathtt{b}\}^*$, amiről tudjuk, hogy reguláris.

2. Tudjuk, hogy az $L_1, L_2, L_3 \subseteq \{\mathtt{0},\mathtt{1}\}^*$ nyelvek közül L_1 és L_3 reguláris. Következik-e ebből, hogy L_2 is reguláris, ha
   (a) $L_3 = L_1 \cap L_2$?
   (b) $L_3 = L_1 \cup L_2$?
   (c) $L_3 = L_1 L_2$?

   Megoldás:
   (a) Nem, pl. $L_1 = \mathtt{0}^*$ és $L_2 = \{\mathtt{0}^n\mathtt{1}^n : n \geq 0\}$ esetén $L_3 = \emptyset$.
   (b) Nem, pl. $L_1 = \mathtt{0}^*\mathtt{1}^*$ és $L_2 = \{\mathtt{0}^n\mathtt{1}^n : n \geq 0\}$ esetén $L_3 = L_1$.
   (c) Nem, pl. ha $L_1 = \emptyset$, akkor függetlenül L_2 választásától $L_3 = \emptyset$.

3. Igaz-e, hogy minden nem üres L nyelvnek
   (a) van olyan $L_r \subseteq L$ nem üres résznyelve, ami reguláris?
   (b) van olyan $L_n \subseteq L$ nem üres résznyelve, ami nem reguláris?

   Megoldás:
   (a) Igen, pl. $L_r = \{s\}$, ahol $s \in L$.
   (b) Nem. Legyen L egy véges (de nem üres) nyelv. Ennek minden részhalmaza is véges, tehát reguláris.

4. Igaz-e, hogy minden véges nyelvhez van 3. osztályba tartozó nyelvtan?

   Megoldás: Igen, hiszen minden ilyen nyelv reguláris, és ezért van reguláris nyelvtana.

   Megjegyzés: nem nehéz felírni egy ilyen nyelvtant, pl. egy $a_1a_2\cdots a_n$ szóhoz egy $S \to a_1A_1$, $A_1 \to a_2A_2$, $\ldots$, $A_{n-2} \to a_{n-1}A_{n-1}$, $A_{n-1} \to a_n$ jó, ahol az A_i-k különböző változók. Ha több szót is akarunk generálni, akkor az S-hez felvesszük mindegyikből az első szabályt.

5. Adjon reguláris nyelvtant ahhoz az $L \subseteq \{\mathtt{0}, \mathtt{1}\}^*$ nyelvhez, amely
   (a) a legalább 1 hosszú, `0`-val kezdődő szavakból áll;
   (b) a legalább 2 hosszú, `00`-val kezdődő szavakból áll;
   (c) a legalább 2 hosszú olyan szavakból áll, melyeknek második karaktere `0`;
   (d) a legalább 2 hosszú `0`-ra végződő szavakból áll!

   Megoldás: Egy-egy lehetséges megoldás, egy kis magyarázattal:

   (a) $S \to \mathtt{0}X \mid \mathtt{0} \qquad X \to \mathtt{0}X \mid \mathtt{1}X \mid \mathtt{0} \mid \mathtt{1}$

   Itt X-ből minden nem üres szó levezethető, ezek elé kell még egy `0`, így a nyelv minden legalább 2 hosszú szava előállítható. Az egy hosszú `0` szó előállításáról gondoskodik a 2. szabály.

   (b) Arra kell figyelni, hogy egyetlen reguláris szabállyal nem állítható elő a `00`, ezt két lépésben kell megtenni. A szónak vagy itt van vége, vagy az előző X változót átemelhetjük ide is:
   $S \to \mathtt{0}E \qquad E \to \mathtt{0}X \mid \mathtt{0} \quad X \to \mathtt{0}X \mid \mathtt{1}X \mid \mathtt{0} \mid \mathtt{1}$.

   (c) Az előzőhöz hasonló, de az első karakter lehet `1` is:
   $S \to \mathtt{0}E \mid \mathtt{1}E \qquad E \to \mathtt{0}X \mid \mathtt{0} \qquad X \to \mathtt{0}X \mid \mathtt{1}X \mid \mathtt{0} \mid \mathtt{1}$.

   (d) Most minden lépésben lehetőség kell legyen `0`-t vagy `1`-t is generálni, kivétel az utolsó lépés, ami `0` lesz. Figyelni kell még arra, hogy legkorábban a második karakternél legyen vége:
   $S \to \mathtt{0}S \mid \mathtt{1}S \mid \mathtt{0}N \mid \mathtt{1}N \quad N \to \mathtt{0}$.

6. Legyen a nyelvtan

$$S \to \mathtt{a}S \mid \mathtt{b}S \mid \mathtt{c}S \mid \mathtt{c}A, \quad A \to \mathtt{c}B \mid \mathtt{c}, \quad B \to \mathtt{a}B \mid \mathtt{b}B \mid \mathtt{c}B \mid \mathtt{a} \mid \mathtt{b} \mid \mathtt{c}$$

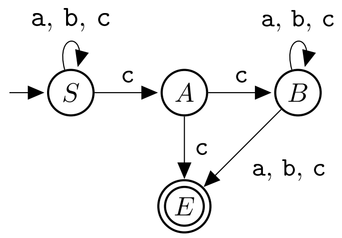
(a) Mi a generált nyelv?

(b) Készítse el a nyelvtanból a tanult módon a megfelelő véges automatát!

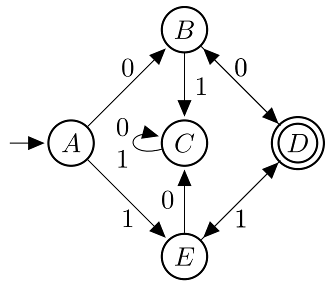
Megoldás:

(a) A nyelv a $\mathtt{cc}$ részszót tartalmazó szavakból áll. Mielőtt az S változót az A felváltja, tetszőleges $\{\mathtt{a}, \mathtt{b}, \mathtt{c}\}^*$-beli szót megkaphatunk, ezt az A-ra váltáskor egy $\mathtt{c}$, majd az A-ból egy újabb $\mathtt{c}$ követi. A szó vagy itt ér véget, vagy a B-ből levezethetőkkel folytatódik. A B változóból pedig a nem üres szavak mindegyike levezethető. Tehát a levezetett szavakban szerepel a $\mathtt{cc}$, de előtte és utána bármi állhat.

(b)



7. Az alábbi automatából a tanult módon készítse el a megfelelő nyelvtant!



Megoldás:

$$A \to \mathtt{0}B \mid \mathtt{1}E \qquad B \to \mathtt{0}D \mid \mathtt{1}C \mid \mathtt{0} \qquad C \to \mathtt{0}C \mid \mathtt{1}C \qquad D \to \mathtt{0}B \mid \mathtt{1}E \qquad E \to \mathtt{0}C \mid \mathtt{1}D \mid \mathtt{1}$$

8. Adjon reguláris nyelvtant ahhoz az $L \subseteq \{\mathtt{0}, \mathtt{1}\}^*$ nyelvhez, amely a páros sok $\mathtt{0}$-t és páratlan sok $\mathtt{1}$-et tartalmazó szavakból áll!

Megoldás: A szokott automatából egyszerűen kapható, de közvetlenül is felírható: Az A_{ij} változó (állapot) felel meg annak, hogy a szóban a $\mathtt{0}$-k paritása i, az $\mathtt{1}$-ek paritása j. Az üres szóban mindkét karakterből nulla darab van, ezért A_{00} a kezdőváltozó. A szónak akkor lehet vége, amikor az A_{01}-be is át tudunk menni.

$$A_{00} \to \mathtt{0}A_{10} \mid \mathtt{1}A_{01} \mid \mathtt{1} \qquad A_{10} \to \mathtt{0}A_{00} \mid \mathtt{1}A_{11} \qquad A_{01} \to \mathtt{0}A_{11} \mid \mathtt{1}A_{00} \qquad A_{11} \to \mathtt{0}A_{01} \mid \mathtt{0} \mid \mathtt{1}A_{10}$$

9. Adjon meg egy CF nyelvtant, amely az $\{\mathtt{a}^k\mathtt{b}^n\mathtt{c}^m : k, n, m \geq 1,\ k = n \text{ vagy } k = m\}$ nyelvet generálja!

Megoldás: Kihasználjuk, hogy $L = L_{ab} \cup L_{ac}$, ahol $L_{ab} = \{\mathtt{a}^k\mathtt{b}^n\mathtt{c}^m : k, n, m \geq 1,\ k = n\}$ és $L_{ac} = \{\mathtt{a}^k\mathtt{b}^n\mathtt{c}^m : k, n, m \geq 1,\ k = m\}$. Ezeket az S_{ab}, illetve az S_{ac} változók generálják, a teljes nyelvtan:

$$S \to S_{ab} \mid S_{ac}$$
$$S_{ab} \to XC \qquad X \to \mathtt{a}X\mathtt{b} \mid \mathtt{ab} \qquad C \to \mathtt{c}C \mid \mathtt{c}$$
$$S_{ac} \to \mathtt{a}S_{ac}\mathtt{c} \mid \mathtt{a}B\mathtt{c} \qquad B \to \mathtt{b}B \mid \mathtt{b}$$

Lényegében az L_{ab} nyelv is két részből van generálva, az X változóból kapjuk az $\{\mathtt{a}^k\mathtt{b}^n : k, n \geq 1,\ k = n\}$ nyelvet, ehhez van hozzáfűzve a C-ből megkapható $\{\mathtt{c}^m : m \geq 1\}$. Hasonló történik az L_{ac} nyelv esetében is.

10. Adjon minél magasabb osztályú nyelvtant, amely az $\{\mathtt{a}^k\mathtt{b}^n\mathtt{c}^m : k \geq 1, m \geq 1, n = k+m\}$ nyelvet generálja!

Megoldás: Példa egy 2. osztályú (CF) nyelvtanra, mely azon az észrevételen alapszik, hogy a nyelv szavai $\mathtt{a}^k\mathtt{b}^k\mathtt{b}^m\mathtt{c}^m$ alakba írhatók, a két rész meg egymástól függtlenül generálható:

$$S \to S_aS_c \qquad S_a \to \mathtt{a}S_a\mathtt{b} \mid \mathtt{ab} \qquad S_c \to \mathtt{b}S_c\mathtt{c} \mid \mathtt{bc}.$$

Azt is megmutatjuk, hogy 3. osztályú nyelvtan nincs a nyelvhez, azaz a nyelv nem reguláris. Tegyük fel, hogy reguláris. Ekkor igaz rá a pumpálási lemma, legyen $p > 0$ a pumpálási hossz. Válasszuk pl. az $\mathtt{a}^p\mathtt{b}^{p+1}\mathtt{c}$ szót. Ez benne van a nyelvben ($k = p, m = 1, n = p+1$), és a hossza is megfelelő ($2p + 2 \geq p$). A szó pumpálása során az $\mathtt{a}$ betűk száma változik csak és ezért a $k + m = n$ egyenlőség elromlik, azaz a szó kikerül a nyelvből, ami ellentmond a pumpálási lemmának, azaz a nyelv nem reguláris.

11. Adjon minél magasabb osztályú nyelvtant, amely a szabályos zárójelsorozatokat generálja! Az abc két eleme (és).

Megoldás: 2. osztályú (CF) nyelvtanra példák (az egyszerűség kedvéért az üres sorozatot kihagyom a nyelvből, hogy az ε-mentesítés ne takarja el a lényeget):
1. változat Azt használja ki, hogy egy jó sorozat felbontható néhány egymás utáni külső zárójelpárra, és mindegyikben egy-egy jó sorozat lehet:

$$Z \to ZZ \mid (Z) \mid ()$$

Az első szabállyal generálható annyi Z, ahány külső zárójelpár kell. A 2-3. szabályokkal oldhatjuk meg ezek belsejét.
2. változat Az első külső zárójellel kezdi a generálást, ebben és utána jó sorozatnak kell állnia:

$$Z \to (Z)Z \mid ()Z \mid ()$$

3. változat Egy tetszőleges külső zárójelpárral kezd:

$$Z \to Z(Z)Z \mid (Z)Z \mid (Z) \mid ()$$

(Igazából ez az üres szót is generáló $X \to X(X)X \mid \varepsilon$ nyelvtannak az üres szót nem generáló CF változata.)

Még meg kell mutatni, hogy a nyelv nem reguláris. Tegyük fel, hogy az, és $p > 0$ a pumpálási hossza. Vegyük azt a z sorozatot, ami p darab nyitó zárójellel kezdődik és utána p darab csukó zárójel van. Ekkor $z \in L$ és $|z| = 2p \geq p$ teljesül. A z minden, a (reguláris) pumpálási lemmának megfelelő $z = uvw$ felosztásában v valahány, de nem nulla nyitó zárójelből áll. Ezért pl. k=2 esetén a pumpált $uv^2w \notin L$, ami ellentmond a pumpálási lemmának, tehát a nyelv nem reguláris.

12. A tanult módszerrel küszöbölje ki az ε-szabályokat az alábbi nyelvtanokból!
    (a) $S \to S\mathtt{a}S\mathtt{b} \mid \epsilon$
    (b) $S \to ABC, \quad A \to BB \mid \varepsilon, \quad B \to CC \mid \mathtt{a}, \quad C \to AA \mid \mathtt{b}$

Megoldás:

(a) Itt nyilván maga az S az elenyésző változó és mivel ez a kezdőváltozó is, ezért kell egy új kezdőváltozó.

$$S' \to \varepsilon \mid S \qquad S \to S\mathtt{a}S\mathtt{b} \mid \mathtt{a}S\mathtt{b} \mid S\mathtt{ab} \mid \mathtt{ab}$$

(b) Elenyésző változók meghatározása: $E_1 = \{A\}$, $E_2 = \{A, C\}$, $E_3 = \{A, B, C\}$, $E_4 = \{A, B, C, S\} = E$.

Az új nyelvtan:

$$\begin{aligned}
S' &\to \varepsilon | S \\
S &\to ABC \mid BC \mid AC \mid AB \mid A \mid B \mid C \\
A &\to BB \mid B \\
B &\to CC \mid C \mid \mathtt{a} \\
C &\to AA \mid A \mid \mathtt{b}
\end{aligned}$$

13. Adjon meg olyan „majdnem CF" nyelvtant, amiben nincs egyszeres szabály, de az ε-szabályok kiküszöbölésekor keletkezik egyszeres szabály!

    Megoldás: Az előző feladat (b) nyelvtana is jó, de itt egy rövidebb: $S \to AB \qquad A \to \mathtt{a}A \mid \varepsilon \qquad B \to \mathtt{b}$. Ez jó, mert $A \to \varepsilon$ miatt az átalakított nyelvtanban lesz $S \to B$ szabály is.

. . . .