

4. Nyelvtanok, CF nyelvek

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

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

$$\begin{aligned} S &\rightarrow S_{ab} \mid S_{ac} \\ S_{ab} &\rightarrow XC \quad X \rightarrow aXb \mid ab \quad C \rightarrow cC \mid c \\ S_{ac} &\rightarrow aS_{ac}c \mid aBc \quad B \rightarrow bB \mid b \end{aligned}$$

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

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

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

$$S \rightarrow S_a S_c \quad S_a \rightarrow a S_a b \mid ab \quad S_c \rightarrow b S_c c \mid bc.$$

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 $a^p b^{p+1} c$ 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 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.

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

Megoldás: 2. osztályú (CF) nyelvtanra példák (az egyszerűség kedvéért az üres sorozatot kihagyom a nyelvből, hogy az ϵ -mentesítés ne takarja el a lényegét):

1. változat Azt használja ki, hogy egy jó sorozat felbontható néhány egymás utáni külső zárójelpárra, és mindegyikben egy-egy jó sorozat lehet:

$$Z \rightarrow ZZ \mid (Z) \mid ()$$

Az első szabállyal generálható annyi Z , ahány külső zárójelpár kell. A 2-3. szabályokkal oldhatjuk meg ezek belsejét.

2. változat Az első külső zárójellel kezd a generálást, ebben és utána jó sorozatnak kell állnia:

$$Z \rightarrow (Z)Z \mid ()Z \mid ()$$

3. változat Egy tetszőleges külső zárójelpárral kezd:

$$Z \rightarrow Z(Z)Z \mid (Z)Z \mid (Z) \mid ()$$

(Igazából ez az üres szót is generáló $X \rightarrow X(X)X \mid \epsilon$ 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.

4. A tanult módszerrel küszöbölje ki az ϵ -szabályokat az alábbi nyelvtanokból!

$$(a) S \rightarrow SaSb \mid \epsilon$$

$$(b) S \rightarrow ABC, \quad A \rightarrow BB \mid \epsilon, \quad B \rightarrow CC \mid a, \quad C \rightarrow AA \mid b$$

Megoldás:

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

$$S' \rightarrow \varepsilon \mid S \quad S \rightarrow SaSb \mid aSb \mid Sab \mid ab$$

(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:

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

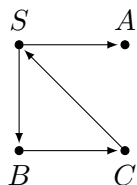
5. Adjon meg olyan „majdnem CF” nyelvtant, amiben nincs egyszeres szabály, de az ε -szabályok kiküszöbölésekor keletkezik egyszeres szabály!

Megoldás: A korábbi, ε -kiküszöbölő feladat (b) nyelvtana is jó, de itt egy rövidebb: $S \rightarrow AB \quad A \rightarrow aA \mid \varepsilon \quad B \rightarrow b$. Ez jó, mert $A \rightarrow \varepsilon$ miatt az átalakított nyelvtanban lesz $S \rightarrow B$ szabály is.

6. A tanult módszerrel szüntesse meg az egyszeres szabályokat a következő nyelvtanban!

$$S \rightarrow A \mid B \quad A \rightarrow aSb \mid a \quad B \rightarrow Sb \mid C \quad C \rightarrow Sa \mid S$$

Megoldás: Az egyszeres szabályok gráfja:



Látszik, hogy az S, B, C változókból mind a 4 csúcs elérhető, A -ból egyetlen más csúcs sem. Az új nyelvtan:

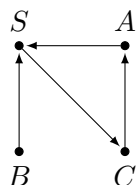
$$S \rightarrow aSb \mid a \mid Sb \mid Sa \quad A \rightarrow aSb \mid a \quad B \rightarrow aSb \mid a \mid Sb \mid Sa \quad C \rightarrow aSb \mid a \mid Sb \mid Sa$$

(A kapott nyelvtanban vannak felesleges szimbólumok, ezek kiküszöbölése további feladat lehet.)

7. A tanult módszerrel alakítsa át a következő nyelvtant olyanra, amelyben már nincsenek egyszeres szabályok és felesleges szimbólumok!

$$S \rightarrow aA \mid Bb \mid C \quad A \rightarrow Ab \mid S \quad B \rightarrow c \mid S \quad C \rightarrow A \mid cc$$

Megoldás: Előbb az egyszeres szabályoktól szabadulunk meg. Ezek gráfja:



Látszik, hogy ebben B -ből minden, az S, A, C változókból B kivételével minden elérhető.

Ennek megfelelően az új nyelvtan:

$$\begin{aligned} S &\rightarrow aA \mid Bb \mid Ab \mid cc \\ A &\rightarrow aA \mid Bb \mid Ab \mid cc \\ C &\rightarrow aA \mid Bb \mid Ab \mid cc \\ B &\rightarrow c \mid aA \mid Bb \mid Ab \mid cc \end{aligned}$$

1. típusú felesleges változók megtalálásához: $T_0 = \{a, b, c\}$ $T_1 = \{a, b, c, S, A, B, C\} = T_2$. Mivel T_2 minden szimbólumot tartalmaz, ezért ilyen típusú felesleges nincs.

2. típus: $S_0 = \{S\}$, $S_1 = \{S, a, A, B, b, c\} = S_2$. Tehát C felesleges, a nélküle kapott nyelvtan:

$$\begin{aligned} S &\rightarrow aA \mid Bb \mid Ab \mid cc \\ A &\rightarrow aA \mid Bb \mid Ab \mid cc \\ B &\rightarrow c \mid aA \mid Bb \mid Ab \mid cc \end{aligned}$$

(Ebben már nincs a definíció értelmében felesleges változó, de kisebb nyelvtan ettől még lehetséges!)

8. A tanult módon szüntesse meg a felesleges szimbólumokat az alábbi környezetfüggetlen nyelvtanban!

$$\begin{aligned} S &\rightarrow AB \mid AC \\ A &\rightarrow aAb \mid bAa \mid a \\ B &\rightarrow bAB \mid aaB \mid BaD \\ C &\rightarrow abCa \mid aBC \mid ba \\ D &\rightarrow bD \mid aCa \end{aligned}$$

Megoldás: Az 1. típusú felesleges szimbólumok (nem tűnnek el) kiszűrése:

$$T_0 = \{a, b\} \quad T_1 = \{a, b, A, C\} \quad T_2 = \{a, b, A, C, S, D\} = T_3,$$

azaz B felesleges, tehát minden B -t tartalmazó szabályt el kell hagyni. Ennek eredménye:

$$\begin{aligned} S &\rightarrow AC \\ A &\rightarrow aAb \mid bAa \mid a \\ C &\rightarrow abCa \mid ba \\ D &\rightarrow bD \mid aCa \end{aligned}$$

A 2. típusú felesleges szimbólumok (nem elérhető) kiszűrése az előbb kapott nyelvtanból:

$$S_0 = \{S\} \quad S_1 = \{S, A, C\} \quad S_2 = \{S, A, C, a, b\} = S_3,$$

azaz D felesleges, tehát minden D -t tartalmazó szabályt el kell hagyni. A felesleges szimbólumok nélküli nyelvtan:

$$\begin{aligned} S &\rightarrow AC \\ A &\rightarrow aAb \mid bAa \mid a \\ C &\rightarrow abCa \mid ba \end{aligned}$$

(Ebből egyszerűbb meghatározni a generált nyelvet is, ugye?)

9. Igazolja, hogy ha egy CF nyelvtanból először az ε -szabályokat küszöböljük ki, utána az egyszeres szabályokat, majd az 1. típusú felesleges szimbólumokat, végül a 2. típusú felesleges szimbólumokat is, akkor a kapott nyelvtanban se ε -szabály (kivéve az esetleges $S \rightarrow \varepsilon$ szabályt), se egyszeres szabály, se felesleges szimbólum nem lesz!

Megoldás: Az egyszeres szabályok kiküszöbölésekor csak a korábban meglevő jobb oldalakat rendeljük más változókhoz is, tehát ha nem volt ε -szabály, akkor ilyen nem is keletkezik.

A felesleges szabályok kiküszöbölésekor csak törölünk néhány szabályt, újat nem adunk hozzá a nyelvtanhoz, ezért nem keletkezik új ε -szabály vagy egyszeres szabály.

Még azt kell látni, hogy a 2. típusú feleslegesek törlésekor nem keletkezik 1. típusú. Ez meg azért van, mert ha a nyelvtan minden változójából levezethető szó, akkor minden, a kezdőváltozóból elérhető változóból indulva, egy szó levezetéseként felhasznált változók is elérhetők a kezdőváltozóból, azaz ezek nem feleslegesek, nem lesznek törölve, így továbbra is levezethető marad a szó.

10. Legyen $\Sigma = \{a, b, c\}$ és L álljon azokból az $a^n b^k c^n$ alakú szavakból, melyekben $0 \leq k \leq 2$, $n \geq 1$, és $n \equiv k \pmod{3}$. Adjon egy, az L nyelvet generáló környezetfüggetlen nyelvtant!

Megoldás: Észrevehetjük, hogy az L 3 nyelv uniója $L_0 = \{a^{3m} c^{3m} : m \geq 1\}$, $L_1 = \{a^{3m+1} b c^{3m+1} : m \geq 0\}$ és $L_2 = \{a^{3m+2} b b c^{3m+2} : m \geq 0\}$.

Egy ennek megfelelő CF nyelvtan a szokásos $\{a^n b^n\}$ nyelvtan mintájára csak itt hármassával rgeneráljuk a karaktereket. Mindig az utolsó szabály gondoskodik a 3-mal való osztás helyes maradékáról.

$$\begin{aligned} S &\rightarrow B_0 \mid B_1 \mid B_2 \\ B_0 &\rightarrow a a b c \\ a a B_0 c c c &\mid a a c c c \\ B_1 &\rightarrow a a a B_1 c c c \mid a b c \\ B_2 &\rightarrow a a a B_2 c c c \mid a a b b c c \end{aligned}$$

11. Legyen $L = \{a^n b^n c^n : n \geq 0\}$. Igazolja, hogy L komplementere környezetfüggetlen!

Megoldás: Az $\bar{L}$ nyelv szavai 4 osztályba sorolhatók:

- nem $a^* b^* c^*$ alakúak, $L_g = \overline{a^* b^* c^*}$,
- $L_{ab} = \{a^i b^j c^k : i \neq j, i, j, k \geq 0\}$,
- $L_{ac} = \{a^i b^j c^k : i \neq k, i, j, k \geq 0\}$,
- $L_{bc} = \{a^i b^j c^k : j \neq k, i, j, k \geq 0\}$.

Egy szó több osztályba is beletartozhat, de $\bar{L} = L_g \cup L_{ab} \cup L_{ac} \cup L_{bc}$ és ezért elég azt megmutatni, hogy ez a 4 nyelv mind CF, mert akkor az uniójuk is.

L_g egy reguláris nyelv komplementere, ezért reguláris, tehát CF is.

Az L_{ab} nyelvet generáló majdnem CF nyelvtan felhasználja, hogy ez két nyelvből áll, az a -kat és b -ket tartalmazó részeknek és c^* -nak az összefűzése. Az első részt lehet úgy generálni, hogy egy ideig egyszerre rakunk le egy a -t és egy b -t, majd egy idő után vagy csak a -kat vagy csak b -ket, így biztosítjuk, hogy valamelyikből több legyen:

$$\begin{aligned} S &\rightarrow S_{ab} C \\ C &\rightarrow c C \mid \varepsilon \\ S_{ab} &\rightarrow a S_{ab} b \mid A \mid B \\ A &\rightarrow a A \mid a \\ B &\rightarrow b B \mid b \end{aligned}$$

Az L_{ac} nyelvtana hasonlóan építhető fel. Most a szélén az a -kat és a c -ket generáljuk együtt egy darabig, utána pedig csak a -kat (legalább egyet) és b -ket, vagy csak c -ket (legalább egyet) és b -ket:

$$\begin{aligned} S &\rightarrow a S c \mid A B \mid B C \\ A &\rightarrow a A \mid a \\ C &\rightarrow c C \mid c \\ B &\rightarrow b B \mid \varepsilon \end{aligned}$$

Az L_{bc} az L_{ab} esethez hasonló, csak ott az elején jön a tetszőleges számú karakter:

$$\begin{aligned} S &\rightarrow AS_{bc} \\ A &\rightarrow \mathbf{a}A \mid \varepsilon \\ S_{bc} &\rightarrow \mathbf{b}S_{bc}\mathbf{c} \mid B \mid C \\ B &\rightarrow \mathbf{b}B \mid \mathbf{b} \\ C &\rightarrow \mathbf{c}C \mid \mathbf{c} \end{aligned}$$

Veremautomatával is egyszerű: pl az L_{ab} nyelvhez amíg az $\mathbf{a}$ -k jönnek, belerakunk egy-egy karaktert a verembe, a $\mathbf{b}$ -knél meg (egy új állapotban) kiszedegetjük. Ha az első $\mathbf{c}$ -nél még van berakott karakter a veremben, az jó, arra persze figyelni kell, hogy után már csak $\mathbf{c}$ -ket olvashatunk. Ha pedig még $\mathbf{b}$ -t olvasunk, amikor a verem alját jelző szimbólumhoz érünk az jó (de ilyenkor is végig kell menni a bemeneten, hogy ellenőrizzük, a $\mathbf{b}$ -k után már csak csupa $\mathbf{c}$ jön).

.....