5. CF nyelvek, pumpálás

1. A tanult módon készítsen üres veremmel elfogadó veremautomatát a következő nyelvtanhoz! Mi a nyelvtan által generált nyelv?

$$S \to XY \qquad X \to XV \mid \mathtt{b} \qquad Y \to UY \mid \mathtt{a} \qquad U \to \mathtt{a} \qquad V \to \mathtt{bb}$$

Megoldás: A veremautomatának egyetlen q állapota lesz, ami egyben a kezdőállapot is. A veremben kezdetben bent levő szimbólum feleljen meg az S változónak Az átmenetek:

$$\delta(q,\varepsilon,S) = \{(q,XY)\} \qquad \delta(q,\varepsilon,X) = \{(q,XV),(q,\mathtt{b})\} \qquad \delta(q,\varepsilon,Y) = \{(q,UY),(q,\mathtt{a})\}$$
$$\delta(q,\varepsilon,U) = \{(q,\mathtt{a})\} \qquad \delta(q,\varepsilon,V) = \{(q,\mathtt{bb})\}$$
$$\delta(q,\mathtt{a},\mathtt{a}) = \{(q,\varepsilon)\} \qquad \delta(q,\mathtt{b},\mathtt{b}) = \{(q,\varepsilon)\}$$

A generált nyelvhez előbb vegyük észre, hogy az Y-ból megkapható szavak halmaza az $\mathtt{aa}^*$, X-ből pedig a páratlan hosszú b-sorozatok kaphatók meg. Mivel a teljes nyelvtan által generált nyelv az előző kettő konkatenáltja, ezért $L = \{\mathtt{b}^{2n-1}\mathtt{a}^k : n, k \geq 1\}$.

2. Környezetfüggetlenek-e az alábbi nyelvek?
   (a) $L_a = \{\mathtt{a}^n\mathtt{b}^k\mathtt{a}^k\mathtt{b}^n : k, n \geq 1\}$
   (b) $L_b = \{\mathtt{a}^n\mathtt{b}^k\mathtt{a}^n\mathtt{b}^k : k, n \geq 1\}$
   (c) $L_c = \{\mathtt{a}^m\mathtt{b}^n : 1 \leq m \leq n \leq 2m\}$
   (d) $L_d = \{\mathtt{0}^{n!} : n \geq 1\}$
   (e) $L_e = \{\mathtt{a}^{i_1}\mathtt{b}^{i_1}\mathtt{a}^{i_2}\mathtt{b}^{i_2}\cdots\mathtt{a}^{i_k}\mathtt{b}^{i_k} : k \geq 0,\ i_1, i_2, \ldots, i_k \geq 1\}$

Megoldás:

(a) Igen. Ezt meg lehet mutatni nyelvtannal és veremautomatával is.
Egy lehetséges nyelvtan, ami arra a gondolatra épül, hogy az 1. és 4. blokkot együtt generáljuk a kezdőváltozóból, majd egy új változó segítségével a középső két blokkot: $S \to \mathtt{a}S\mathtt{b} \mid \mathtt{a}T\mathtt{b} \qquad T \to \mathtt{b}T\mathtt{a} \mid \mathtt{ba}$

Egy megfelelő veremautomata vázlata: A kezdő a betű olvasásakor egy q_{a1} állapotba lép, és közben egy A-val bővíti a vermet. Ebben az állapotban a további a betűkre is ugyanezt csinálja. Az első b-re átlép egy q_{b1} állapotba és egy B-vel bővíti a vermet. Ugyanez történik a további b betűknél. Innen egy a betűre átlép q_{a2}-be, és minden a hatására kivesz a veremből egy-egy B-t. Egy b-vel q_{b2}-be kerül és innentől minden b hatására kivesz egy-egy A-t. Kell még egy $(q_{b2}, \varepsilon, Z_0) \to (q_f, Z_0)$ átmenet, amivel az elfogadó q_f állapotba léphet.

(b) Nem. Tegyük fel, hogy CF és legyen $p > 0$ a pumpálási hossz. Válasszuk a $z = a^pb^pa^pb^p \in L_b$ szót. Világos, hogy $|z| = 4p \geq p$. Vegyük ennek egy tetszőleges, a pumpálási lemma szerinti $z = uvwxy$ felosztását.

Ha v-ben kétféle betű is van, akkor az ismétlésekor a blokkok száma nő, a pumpált változat nem lesz a nyelvben. Ugyanez igaz, amikor x-ben van kétféle betű. Tehát mind a két részszó homogén.

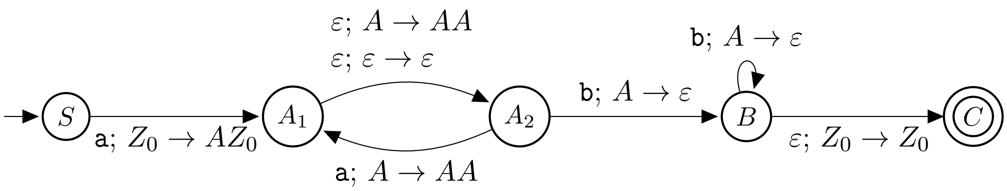
Ha azonos blokkon belül vannak, akkor pumpáláskor ennek a blokknak a hossza változik, de a párjáé nem, ez is kivezet a nyelvből. Ugyanez történik, ha v vagy x az üres szó. (Tudjuk, hogy mindkettő nem lehet üres.)

Még az az eset maradt, hogy v és x különböző blokkokban van és egyik sem üres. Mivel a lemma szerint $|vwx| \leq p$, ez csak két szomszédos blokk lehet. Pummpáláskor ezek hossza változik, de a párjaiké nem, ez sem lesz benne a nyelvben. Azaz ellentmondásra jutottunk, tehát L_b nem CF.

(c) Igen.
A nyelvtan ötlete lehet, hogy szokás szerint a két szélén generáljuk az a és b betűket, de most egy a-val együtt vagy 1 vagy 2 b betű keletkezik:

$$S \to \mathtt{a}S\mathtt{b} \mid \mathtt{a}S\mathtt{bb} \mid \mathtt{ab} \mid \mathtt{abb}$$

A veremautomata hasonló elven alapulhat: pl. minden a betűnél berakunk egy vagy két A szimbólumot a verembe (az utóbbihoz két lépés kell!), az első b betűvel átlépünk egy új állapotba, ahonnan kezdve minden b betűnél egy A-t szedünk ki. Egy lehetséges megvalósítás:



(d) Nem. Tegyük fel, hogy CF és legyen $p > 0$ a pumpálási hossz. Válasszuk a $z = 0^{p!} \in L_d$ szót. Világos, hogy $|z| = p! \geq p$. Vegyük ennek egy tetszőleges, a pumpálási lemma szerinti $z = uvwxy$ felosztását. Ha $t = |vx|$, akkor $1 \leq t \leq p$. Pumpáláskor a szó hossza $|uv^kwx^ky| = p! + (k-1)t$ lesz. Ez pl. $k = 2$ esetén $p! + t \leq p! + p < (p+1)!$, tehát nem $n!$ alakú, $uv^2wx^2y \notin L_d$, ami ellentmond a pumpálási lemmának, ezért L_d nem CF nyelv.

(e) Igen. Vegyük észre, hogy $L_e = L^*$, ahol $L = \{\mathtt{a}^n\mathtt{b}^n : n \geq 1\}$. Mivel L CF és a CF nyelvek zártak a tranzitív lezárásra, ezért L_e is CF.
(Természetesen lehet nyelvtant és veremautomatát is adni rá.)

3. Legyen $L = \{\mathtt{a}^{3i}\mathtt{b}^i\mathtt{c}^{2i} : |i \geq 1\}$. Bizonyítsa be, hogy ez az L nyelv nem környezetfüggetlen!

Megoldás: Tegyük fel, hogy CF és legyen $p > 0$ a pumpálási hossz. Válasszuk pl. a $z = \mathtt{a}^{3p}\mathtt{b}^p\mathtt{c}^{2p} \in L$ szót. Ennek hossza $|z| = 6p \geq p$, tehát a szó eleget tesz a pumpálási lemma feltételeinek. Vegyük ennek egy tetszőleges $z = uvwxy$ felosztását.

Ha v nem csak egyféle betűből áll, akkor pl. az $uvvwxxy$ szó nem 3, hanem több blokkból áll, ezért ez nincs a nyelvben. Ugyanez igaz x-re is.

Ha v és x is ugyanabban a blokkban van (az egyik akár üres is lehet), akkor pumpáláskor ennek a blokknak a hossza változik, míg a másik kettővé nem, a szó kikerül a nyelvből. Ha v és x különböző blokkokban vannak (és egyik sem üres), akkor is csak két blokk lesz érintve a pumpáláskor, a harmadik hossza nem változik, így sem maradunk a nyelvben.

Tehát minden felosztás ellentmond a pumpálási lemmának, a nyelv nem lehet CF.

4. Legyen $L = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k : 1 \leq i, j, k, \text{ és } i < j, k < j\}$ Bizonyítsa be, hogy ez az L nyelv nem környezetfüggetlen!

Megoldás: Tegyük fel, hogy CF és legyen $p > 0$ a pumpálási hossz. Válasszuk pl. a $z = \mathtt{a}^p\mathtt{b}^{p+1}\mathtt{c}^p \in L$ szót. Ennek hossza $|z| = 3p + 1 \geq p$, tehát a szó eleget tesz a pumpálási lemma feltételeinek. Vegyük ennek egy tetszőleges $z = uvwxy$ felosztását.

Ha v nem csak egyféle betűből áll, akkor pl. az $uvvwxxy$ szó nem 3, hanem több blokkból áll, ezért ez nincs a nyelvben. Ugyanez igaz x-re is.

Ha v és x is ugyanabban a blokkban van (az egyik akár üres is lehet), akkor az a- és a c-blokk esetén felfelé pumpáláskor a megfelelő blokk hossza eléri vagy meghaladja a b-blokk hosszát, a szó nem lesz a nyelvben. Ha a b-blokkban van mindkét részszó, akkor a lefelé pumpálás viszi ki a nyelvből.

Ha v és x különböző blokkokban vannak (és egyik sem üres), akkor, mivel $|vwx| \leq p$ az egyikük a b-blokkban van. Megint a lefelé pumpálás visz ki a nyelvből, hiszen a harmadik, nem érintett blokk hosszánál kevesebb b lesz. Tehát a szó minden felosztása ellentmond a pumpálási lemmának, a nyelv nem lehet CF.

5. Legyen $L = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k : i < j < k\}$. Igaz-e, hogy ez a nyelv
   (a) reguláris
   (b) környezetfüggetlen
   (c) nem környezetfüggetlen ?

Megoldás: Megmutatjuk, hogy (c) igaz, és akkor persze (a) és (b) nem igaz.

Tegyük fel, hogy L egy CF nyelv és legyen $p > 0$ a pumpálási hossza.

Válasszuk pl. a $z = \mathtt{a}^p\mathtt{b}^{p+1}\mathtt{c}^{p+2} \in L$ szót. Ennek hossza $|z| = 3p+3 \geq p$, tehát a szó eleget tesz a pumpálási lemma feltételeinek. Vegyük ennek egy tetszőleges $z = uvwxy$ felosztását.

Ha v nem csak egyféle betűből áll, akkor pl. az $uvvwxxy$ szó nem 3, hanem több blokkból áll, ezért ez nincs a nyelvben. Ugyanez igaz x-re is.

Ha v és x is ugyanabban a blokkban van (az egyik akár üres is lehet), akkor az a- és a b-blokk esetén felfelé pumpáláskor a megfelelő blokk hossza eléri vagy meghaladja az utána levő blokk hosszát, a szó nem lesz a nyelvben. Ha a c-blokkban van mindkét részszó, akkor a lefelé pumpálás viszi ki a nyelvből.

Ha v és x különböző blokkokban vannak (és egyik sem üres), akkor, mivel $|vwx| \leq p$, vagy egyikük sincs az a-blokkban és ekkor a lefelé pumpáláskor túl kevés b lesz, vagy egyikük sincs a c-blokkban, amikor meg a felfelé pumpáláskor túl sok b lesz. Tehát a szó minden felosztása ellentmond a pumpálási lemmának, a nyelv nem lehet CF.

6. Adott $L_1, L_2 \subseteq \{\mathtt{a}, \mathtt{b}\}^*$ nyelvekhez legyen $L_\oplus$ az azokból a szavakból álló nyelv, melyek az L_1 és L_2 nyelvek közül pontosan az egyikben vannak benne. Igazolja, hogy van olyan L_1 és L_2 környezetfüggetlen nyelv, melyre
   (a) $L_\oplus$ is környezetfüggetlen!
   (b) $L_\oplus$ nem környezetfüggetlen!

Megoldás:

(a) Legyen pl. L_1 tetszőleges CF nyelv és $L_2 = L_1$. Ekkor $L_\oplus = \emptyset$, ami környezetfüggetlen (reguláris is). Egy másik egyszerű példa: $L_1 = \emptyset$, L_2 tetszőleges CF nyelv. Ekkor $L_\oplus = L_2$.

(b) Legyen $L_1 = \{\mathtt{a}, \mathtt{b}\}^*$. Ekkor $L_\oplus = \overline{L_2}$. Tehát ha L_2 egy olyan CF nyelv, aminek komplementere nem CF (tudjuk, hogy van ilyen), akkor $L_\oplus$ nem környezetfüggetlen.

7. Lehetséges-e, hogy ha az $L_1, L_2 \subseteq \Sigma^*$ nyelvek környezetfüggetlenek, akkor az $L_1 \cup L_2$ és az $L_1 \cap L_2$ nyelvek közül
   (a) pontosan az egyik környezetfüggetlen?
   (b) mindkettő környezetfüggetlen?
   (c) egyik sem környezetfüggetlen?

Megoldás: Tudjuk, hogy a CF nyelvek zártak az unióra, de a metszetre nem. Ennek alapján

(a) Lehet: L_1 és L_2 legyen olyan CF, amiknek metszete nem CF.

(b) Lehet, pl. $L_1 = L_2$ esetén az uniójuk és a metszetük is az L_1, tehát CF.

(c) Nem lehet, hiszen az unió mindig CF.