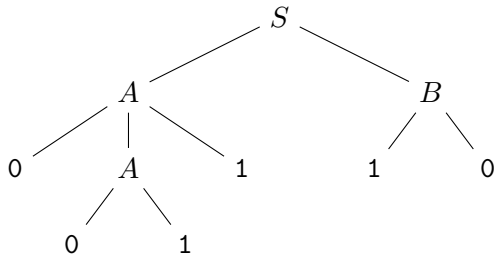


Veremautomaták, környezetfüggetlen nyelvek

1. Legyen a nyelvtan: $S \rightarrow AB \quad A \rightarrow 0A1 \mid 01 \quad B \rightarrow 1B0 \mid 10$

- (a) Adjon meg egy levezetési fát és egy bal-levezetést a 001110 szóhoz!
- (b) Határozza meg a nyelvtan által generált nyelvet!

Megoldás: (a)



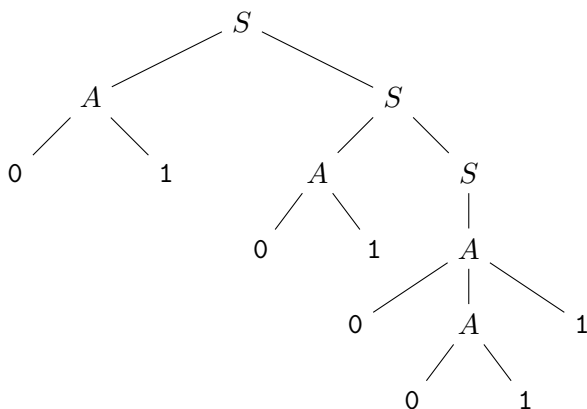
A fából kiolvasott bal-levezetés: $S \Rightarrow \underline{A}B \Rightarrow \underline{0A}1B \Rightarrow 00\underline{11}B \Rightarrow 0011\underline{10}$

(b) A -ból az $L_A = \{0^n 1^n : n \geq 1\}$ nyelv vezethető le, B -ből pedig az $L_B = \{1^k 0^k : k \geq 1\}$, ezért S -ből minden, ami egy-egy ilyen szó összefűzésével megkapható, azaz $L = \{0^n 1^{n+k} 0^k\}$.

2. Legyen $S \rightarrow AS \mid A \quad A \rightarrow 0A1 \mid 01$

- (a) Adjon meg egy levezetési fát és egy bal-levezetést a 01010011 szóhoz!
- (b) Határozza meg a nyelvtan által generált nyelvet!

Megoldás: (a)



A fából kiolvasott bal-levezetés: $S \Rightarrow \underline{A}S \Rightarrow 01\underline{S} \Rightarrow 01\underline{A}S \Rightarrow 0101\underline{S} \Rightarrow 0101\underline{A} \Rightarrow 01010\underline{A}1 \Rightarrow 01010011$

(b) A -ból az $L_A = \{0^n 1^n : n \geq 1\}$ nyelv vezethető le, S -ből pedig azok a szavak, amelyek néhány L_A -beli összefűzésével keletkeznek, $L = \{0^{n_1} 1^{n_1} 0^{n_2} 1^{n_2} \dots 0^{n_k} 1^{n_k} : k \geq 1, n_1, n_2, \dots, n_k \geq 1\}$.

3. Adjon környezetfüggetlen nyelvtant az alábbi nyelvekre! Egyértelműek a megadott nyelvtanok?

- (a) $L = \{a^n b^{n+1} : n \geq 0\}$
- (b) $L = \{a^n b^{2n} : n \geq 0\}$
- (c) Az $\{a, b\}$ feletti palindromok nyelve
- (d) $L = \{a^n b^k c^\ell : n = k \text{ vagy } n = \ell\}$

Megoldás: Mindegyiknél több jó megoldás is van, pl.

(a) $S \rightarrow T\mathbf{b}, \quad T \rightarrow \mathbf{a}T\mathbf{b} \mid \varepsilon$

Ez egyértelmű nyelvtan, mert minden levezetés az első szabállyal kezdődik, utána ahhoz, hogy az $a^n b^{n+1}$ szót kapjuk n -szer kell alkalmaznunk a 2. és a végén egyszer a 3. szabályt.

Egy másik lehetőség például, ha lényegében az $a^n b^n$ nyelvet generáljuk, és a végén rakjuk a szó "közepére" az extra b betűt, azaz $S \rightarrow \mathbf{a}S\mathbf{b} \mid \mathbf{b}$ (ez a nyelvtan is egyértelmű).

(b) Az ötlet, hogy most egy lépésben egy a betűt teszünk előre és két b -t hátra, $S \rightarrow aSbb \mid \varepsilon$
 Világos, hogy az így kapott nyelvtan jó. Ez a nyelvtan egyértelmű, mert n -szer kell alkalmazni az 1., és egyszer a 2. szabályt.

(c) Vegyük észre, hogy a legalább 2 hosszú palindromok úgy épülnek fel, hogy az első és az utolsó karakterük megegyezik, és köztük egy (esetleg nulla hosszú) palindrom van. Ennek alapján egy lehetséges nyelvtan:
 $S \rightarrow aSa \mid bSb \mid a \mid b \mid \varepsilon$

A nyelvtan egyértelmű, hiszen a szótól függően mindig csak az egyik szabály alkalmazható.

(d) L két nyelv uniója, az egyik, amikor $n = k$ és ℓ tetszőleges, a másik amikor $n = \ell$ és k tetszőleges. A nyelvtan készülhet úgy, hogy erre a két nyelvre megadunk egy-egy nyelvtant, és ezeket "összerakjuk". Az X változóból lesz levezethető az első, az Y -ből a második nyelv. Az X -ből levezethető szavak eleje $a^n b^n$, amit mondjuk az A változóból kapunk (erre már láttunk nyelvtant), ezt kell kiegészíteni még tetszőleges számú c -vel, amit meg a C változó old meg. Az Y -ből levezethető szavaknak meg a széle készülhet az $a^n c^n$ mintájára, a közepére kell tetszőleges számú b , amit az előzőhöz hasonlóan oldunk meg egy B változóval:
 $S \rightarrow X \mid Y, X \rightarrow AC, A \rightarrow aAb \mid \varepsilon, C \rightarrow cC \mid \varepsilon, Y \rightarrow aYc \mid B, B \rightarrow bB \mid \varepsilon$

(Az X változó el is hagyható, ha a hozzá tartozó egyetlen szabály jobb oldalát írjuk a helyébe.)

A nyelvtan nem egyértelmű, mert pl. az abc szó megkapható az X és az Y változóból is. (Minden $n = k = \ell$ típusú szó kétféleképpen is levezethető, és ez nem is kerülhető el, egy tétel szerint erre a nyelvre nincs is egyértelmű nyelvtan.)

4. Egyértelműek-e az alábbi nyelvtanok? Egyértelműek-e az általuk generált nyelvek?

(a) $S \rightarrow aSa \mid bSb \mid aa \mid bb \mid a \mid b$

(b) $S \rightarrow TT \mid U \quad T \rightarrow 0T \mid T0 \mid \# \quad U \rightarrow 0U00 \mid \#$

Megoldás: Az (a) nyelvtan egyértelmű. Ehhez azt kell észrevenni, hogy a szó következő karaktere, és hogy hány betűt kell még generálnunk egyértelműen meghatározza, melyik szabályt kell alkalmazni.

Mivel a nyelvtan egyértelmű, azért a generált nyelv (ami a nem üres palindromok nyelve) is egyértelmű.

(b) Próbáljuk megérteni, mit csinál a nyelvtan! A T változóból levezethető szavak a $0^k \# 0^n$ alakúak ($k, n \geq 0$), TT -ből a $0^k \# 0^n 0^r \# 0^s$ alakúak, azaz a $0^k \# 0^t \# 0^s$ alakúak. Az U -ből a $0^k \# 0^{2k}$ alakúak ($k \geq 0$). Tehát a nyelv $L = \{0^k \# 0^t \# 0^s : k, t, s \geq 0\} \cup \{0^k \# 0^{2k} : k \geq 0\}$. Az, hogy a levezetendő szóban egy vagy két $\#$ szerepel, meghatározza, hogy az első vagy a második szabállyal kell indulni, tehát ez a lépés egyértelmű. Nézzük mi van, ha az első $S \rightarrow TT$ szabállyal kezdünk. Például a $0\#0\#0$ szót megkaphatjuk úgy, hogy az első T -ből a $0\#0$ szót vezetjük le, és a többit a másodikból, de úgy is, hogy a $0\#$ részt vezetjük csak le az első T -ből és a többit a másodikból. Tehát a nyelvtan nem egyértelmű. (Ha a szó az első típusú és $t > 0$, akkor a levezetése nem egyértelmű. Igazából már a T -ből való levezetés sem egyértelmű: attól függően, hogy milyen sorrendben alkalmazzuk a $0T$ és a $T0$ helyettesítéseket, más-más levezetési fát kapunk.)

A nyelv egyértelműségéhez mutatnunk kell hozzá egy olyan nyelvtant, ami egyértelmű. Nem nehéz ilyen csinálni a nyelv ismeretében: az U jó, ahogy van, a T változót cseréljük le például így $S \rightarrow N\#N\#N \mid U, N \rightarrow 0N \mid \varepsilon, U \rightarrow 0U00 \mid \varepsilon$

A levezetendő szó alapján kell választanunk a kezdő szabályt. Az U szabályai nem változtak, N -ből tetszőleges 0^k levezethető, méghozzá egyértelműen.

5. Legyen az ábécé $\Sigma = \{0, 1\}$, a veremautomata állapotai $Q = \{A, B, C\}$, amiből A a kezdő állapot és C az egyetlen elfogadó állapot, Z a verem kezdő szimbóluma. A veremautomata állapotátmeneti függvénye: $\delta(A, 0, \varepsilon) = \{(A, a)\}$, $\delta(A, 1, \varepsilon) = \{(A, b)\}$, $\delta(A, \varepsilon, \varepsilon) = \{(B, \varepsilon)\}$, $\delta(B, 0, a) = \{(B, \varepsilon)\}$, $\delta(B, 1, b) = \{(B, \varepsilon)\}$, $\delta(B, \varepsilon, Z) = \{(C, \varepsilon)\}$.

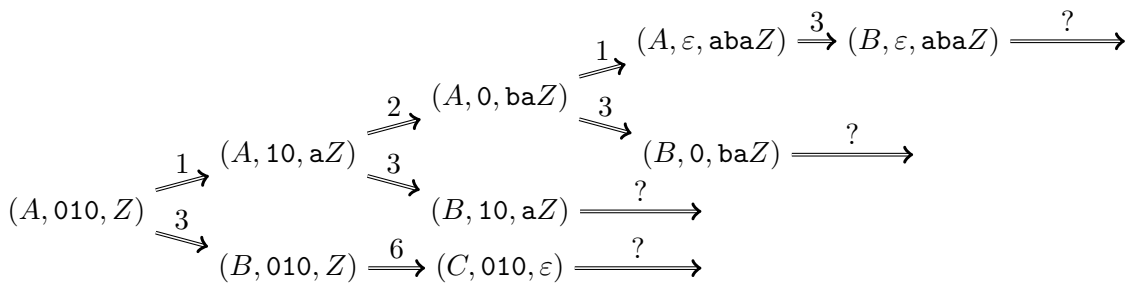
(a) Adja meg a lehetséges számításokat a 010 szó esetén!

(b) Elfogadja az automata a 0110 szót?

(c) Mi az automata által elfogadott nyelv?

(d) Adjon meg ehhez a nyelvhez egy környezetfüggetlen nyelvtant!

Megoldás: Az alábbi számítási fából látszik, hogy a veremautomata nem fogadja el a 010 szót.



(b) A 0110 szót elfogadja, mert ha az első két lépésben az A -ban maradunk, a veremben (felülről nézve) baZ lesz. Ha itt átmegyünk a B állapotba, akkor a 3. karakterre a b betűt, a negyedikre meg az a betűt kiszedhetjük a veremből. Amikor a bemeneti szó elfogyott, akkor a veremben csak a Z van, és ekkor átlépünk a C elfogadó állapotba.

(c) A veremautomata a páros hosszú palindromokat fogadja el. A (b) pont mintájára könnyű látni, hogy ezeket tényleg elfogadja. Másrészt, egy elfogadó számítás során amíg az A állapotban van, a bemeneti $w = a_1a_2 \dots a_n$ szó elejének megfelelően rak a és b betűket a verembe. Mivel a C -ben már nem olvas, a B állapotban kell, hogy a szó további részét megvizsgálja. Ha ide k darab karakter A -beli feldolgozása után kerül, akkor abban az esetben nem akad el, ha $a_{k+1} = a_k$, $a_{k+2} = a_{k-1}$, stb. Ahhoz, hogy a verem pont akkor ürüljön ki, amikor a szó végére ér az kell, hogy $n = 2k$ teljesüljön.

Vegyük észre, hogy ez a veremautomata az üres szót is elfogadja (ami jó, mert az is páros hosszú palindrom).

(d) Egy lehetőség: $S \rightarrow aSa \mid bSb \mid \varepsilon$

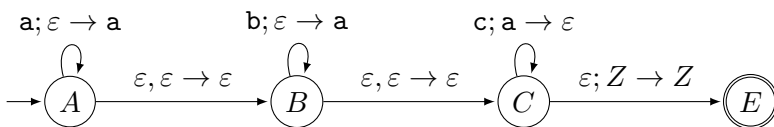
6. Készítsen veremautomatát a palindromok nyelvéhez!

Megoldás: Az előzőt annyiban kell módosítani, hogy lehessen egy tetszőleges "középső" karakter is. Ezt megoldja, ha az A állapotból a B -be egy karakter olvasásával a verem változtatása nélkül is át tudunk lépni, azaz az előző feladat átmeneti szabályainál az első kettőt így módosítjuk: $\delta(A, 0, \varepsilon) = \{(A, a), (B, \varepsilon)\}$, $\delta(A, 1, \varepsilon) = \{(A, b), (B, \varepsilon)\}$,

7. Adjon veremautomatát az alábbi nyelvekhez!

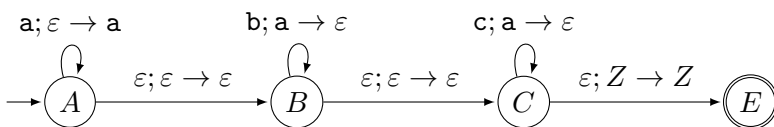
- (a) $L_a = \{a^i b^j c^k : i, j, k \geq 0 \text{ és } i + j = k\}$
- (b) $L_b = \{a^i b^j c^k : i, j, k \geq 0 \text{ és } j + k = i\}$
- (c) $L_c = \{a^i b^j c^k : i, j, k \geq 0 \text{ és } i + k = j\}$

Megoldás: (a) Az ötlet az, hogy amikor a vagy b karakter jön, akkor mondjuk mindig egy a -t rakunk a verembe, a c -kre meg egyet kiszedünk. Akkor fogadunk el, ha a szó végén csak a Z marad a veremben. Persze arra is vigyáznunk kell, hogy b után nem jöhet a , c után meg már csak c jöhet akkor, ha elfogadjuk a szót. Lesz tehát 4 állapot, az A -ban az a -knál, a B -ben a b -knél, a C -ben a c -knél tartunk. Innen léphetünk át az E állapotba, ami az egyetlen elfogadó állapot.

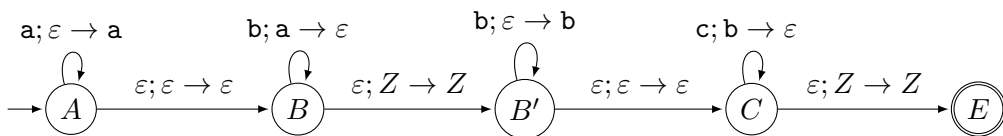


A különböző állapotok közötti $\varepsilon; \varepsilon \rightarrow \varepsilon$ átmenetek biztosítják azt is, hogy az üres szót ($i = j = k = 0$) elfogadja.

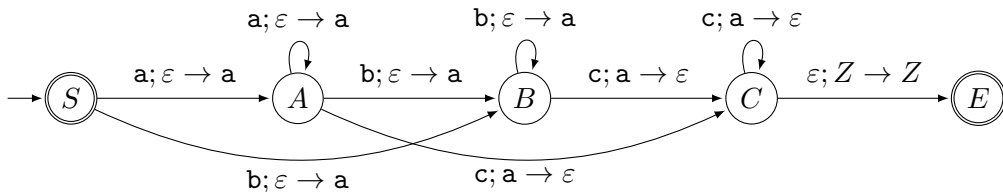
(b) Hasonló az előzőhöz, de most csak az a -kra rakunk be egy-egy a -t a verembe, b -re és c -re is kiveszünk egyet.



(c) Az a -kra megint betehetünk egy-egy a -t a verembe, de a b -knél ki kell ezeket szedni. Ha közben a verem kiürül, akkor onnantól kezdve mondjuk b -ket rakunk be, amiket majd a c -k olvasásakor szedegetünk ki.



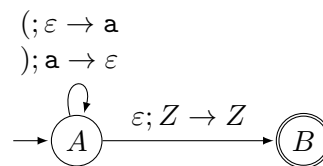
Megjegyzés: a fent megadott veremautomaták mindegyike nemdeterminisztikus az új állapotba vezető átmenetek miatt, de ezekhez a nyelvekhez lehet determinisztikus veremautomatát is adni. Pl. az első esetben:



8. Készítsen veremautomatát a jó zárójelezések nyelvéhez!

Megoldás:

Az ötlet hasonló az előzőekhez, a nyitó zárójelnél berakunk egy a betűt a verembe, a csukónál kiveszünk, különbség, hogy csukó zárójel után is jöhet nyitó, ezért nem kell annyi állapot. Az, hogy egy jó zárójelsorozattal van dolgunk, pontosan azt jelenti, hogy csukó zárójelnél mindig van legalább egy a a veremben de a szó végére egy sem marad.



9. Adjon veremautomatát az alábbi nyelvekhez!

(a) $L_a = \{a^n b^m : 2n = m \geq 1\}$

(b) $L_b = \{a^n b^m : 2n \geq m \geq n \geq 1\}$

Megoldás: Itt most már csak szavakban vázoljuk a veremautomaták működését.

(a) A kezdő állapotból az első a olvasásakor két darab a betűt rak a verembe és átmegy az A állapotba. Itt minden további a betűnél megint két darab a betűt rak a verembe. Az A -ból az első b olvasásakor kivesz a veremből egy a -t és átlép a B állapotba, ahol minden további olvasott b betűnél kivesz egy a -t a veremből. Végül olvasás nélkül a verem alját jelző Z -t látva átlép a C elfogadó állapotba, ahol a számítás véget ér. Világos, hogy ha $m = 2n$, akkor a szó végén az elfogadó állapotba juthatunk. Ha pedig rosszkor váltunk állapotot vagy a szó nem megfelelő, akkor vagy nem jutunk el a C -ig vagy a szó addigra nem ér véget, a számítás mindenképpen elakad, és ezért nem elfogadó.

(b) Kihasználjuk a nemdeterminizmus adta lehetőségeket: az előző automatát úgy módosítjuk, hogy az A állapotban minden a betűre vagy egy vagy kettő a -t rakunk a verembe. Ha egy szó az L_b nyelvben van, akkor lesz hozzá elfogadó számítás (amikor $(2n - m)$ -szer rakunk be egyet és $(m - n)$ -szer kettőt). Ha az m túl kicsi, a verem nem fog kiürülni, ha meg nagy, akkor nem tudunk kivenni belőle, amikor kellene, a számítás elakad.

10. Tekintsük az előadáson szerepelt egyértelmű nyelvtant az aritmetikai kifejezésekre.

(a) Adjon meg az $a + a * a * a + a$ szóhoz egy levezetési fát!

(b) Ha a levezetési fa alapján számoljuk ki a kifejezés értékét, akkor milyen sorrendben végezzük el a műveleteket?

(c) Egészítse ki a nyelvtant úgy, hogy a kivonás is szerepeljen benne!

(d) Mi lesz ennél a nyelvtannál az $a - a + a - a$ levezetési fája? Mi a műveletek sorrendje?

Megoldás: (a)

