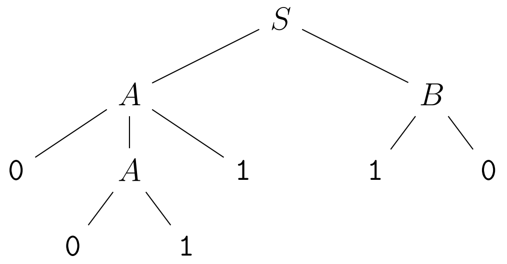
1. Legyen a nyelvtan: $S \to AB$; $A \to \mathtt{0}A\mathtt{1} \mid \mathtt{01}$; $B \to \mathtt{1}B\mathtt{0} \mid \mathtt{10}$

   (a) Adjon meg egy levezetési fát a `001110` szóhoz!

   (b) Határozza meg a nyelvtan által generált nyelvet!

   Megoldás: (a)



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

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

   (a) Az $\{\mathtt{a}, \mathtt{b}\}$ feletti palindromok nyelve

   (b) $L = \{\mathtt{a}^n\mathtt{b}^k\mathtt{c}^\ell : n = k \text{ vagy } n = \ell\}$

   Megoldás: Mindegyiknél több jó megoldás is van, pl.:
   (a) 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: $S \to \mathtt{a}S\mathtt{a} \mid \mathtt{b}S\mathtt{b} \mid \mathtt{a} \mid \mathtt{b} \mid \varepsilon$
   A nyelvtan egyértelmű, hiszen a szótól függően mindig csak az egyik szabály alkalmazható.
   (b) 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-ból a második nyelv. Az X-ből levezethető szavak eleje $\mathtt{a}^n\mathtt{b}^n$, 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-ból levezethető szavaknak meg a széle készülhet az $\mathtt{a}^n\mathtt{c}^n$ mintájára, a közepére kell tetszőleges számú `b`, amit az előzőhöz hasonlóan oldunk meg egy B változóval:
   $S \to X \mid Y$, $X \to AC$, $A \to \mathtt{a}A\mathtt{b} \mid \varepsilon$, $C \to \mathtt{c}C \mid \varepsilon$, $Y \to \mathtt{a}Y\mathtt{c} \mid B$, $B \to \mathtt{b}B \mid \varepsilon$
   (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.)

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

   (a) $S \to \mathtt{a}S\mathtt{a} \mid \mathtt{b}S\mathtt{b} \mid \mathtt{aa} \mid \mathtt{bb} \mid \mathtt{a} \mid \mathtt{b}$

   (b) $S \to TT \mid U \qquad T \to \mathtt{0}T \mid T\mathtt{0} \mid \# \qquad U \to \mathtt{0}U\mathtt{00} \mid \#$

   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 $\mathtt{0}^k\#\mathtt{0}^n$ alakúak ($k, n \geq 0$), TT-ből a $\mathtt{0}^k\#\mathtt{0}^n\mathtt{0}^r\#\mathtt{0}^s$ alakúak, azaz a $\mathtt{0}^k\#\mathtt{0}^t\#\mathtt{0}^s$ alakúak. Az U-ból a $\mathtt{0}^k\#\mathtt{0}^{2k}$ alakúak ($k \geq 0$). Tehát a nyelv: $L = \{\mathtt{0}^k\#\mathtt{0}^t\#\mathtt{0}^s : k, t, s \geq 0\} \cup \{\mathtt{0}^k\#\mathtt{0}^{2k} : k \geq 0\}$. 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 $S \to TT$ szabállyal kezdünk. Például a $\mathtt{0}\#\mathtt{0}\#\mathtt{0}$ szót megkaphatjuk úgy, hogy az első T-ből a $\mathtt{0}\#\mathtt{0}$ szót vezetjük le, és a többit a másodikból, de úgy is, hogy a $\mathtt{0}\#$ 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 $\mathtt{0}T$ és a $T\mathtt{0}$ helyettesítéseket, más-más levezetési fát kapunk.)

A nyelv egyértelműségéhez mutatnunk kell hozzá egy olyan nyelvtant, ami egyértelmű. Nem nehéz ilyet csinálni a nyelv ismeretében: az U jó, ahogy van, a T változót cseréljük le például így $S \to N\#N\#N \mid U, \quad N \to \mathtt{0}N \mid \varepsilon, \quad U \to \mathtt{0}U\mathtt{00} \mid \#$
A levezetendő szó alapján kell választanunk a kezdő szabályt. Az U szabályai nem változtak, N-ből tetszőleges 0^k levezethető, méghozzá egyértelműen.

4. Legyen az ábécé $\Sigma = \{\mathtt{0}, \mathtt{1}\}$, 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: $\delta(A, \mathtt{0}, \varepsilon) = \{(A, a)\}$, $\delta(A, \mathtt{1}, \varepsilon) = \{(A, b)\}$, $\delta(A, \varepsilon, \varepsilon) = \{(B, \varepsilon)\}$, $\delta(B, \mathtt{0}, a) = \{(B, \varepsilon)\}$, $\delta(B, \mathtt{1}, b) = \{(B, \varepsilon)\}$, $\delta(B, \varepsilon, Z) = \{(C, \varepsilon)\}$.
   (a) Adja meg a lehetséges számításokat a $\mathtt{010}$ szó esetén!
   (b) Elfogadja az automata a $\mathtt{0110}$ 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 $\mathtt{010}$ szót.

$$(A,\mathtt{010},Z) \overset{1}{\Longrightarrow} (A,\mathtt{10},\mathtt{a}Z) \overset{2}{\Longrightarrow} (A,\mathtt{0},\mathtt{ba}Z) \overset{1}{\Longrightarrow} (A,\varepsilon,\mathtt{aba}Z) \overset{3}{\Rightarrow} (B,\varepsilon,\mathtt{aba}Z) \overset{?}{\Longrightarrow}$$
$$(A,\mathtt{0},\mathtt{ba}Z) \overset{3}{\Longrightarrow} (B,\mathtt{0},\mathtt{ba}Z) \overset{?}{\Longrightarrow}$$
$$(A,\mathtt{10},\mathtt{a}Z) \overset{3}{\Longrightarrow} (B,\mathtt{10},\mathtt{a}Z) \overset{?}{\Longrightarrow}$$
$$(A,\mathtt{010},Z) \overset{3}{\Longrightarrow} (B,\mathtt{010},Z) \overset{6}{\Longrightarrow} (C,\mathtt{010},\varepsilon) \overset{?}{\Longrightarrow}$$

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

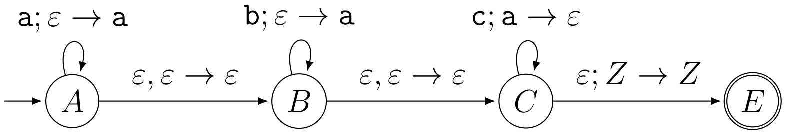
(c) A veremautomata a páros hosszú palindromokat fogadja el. A (b) pont mintájára könnyű látni, hogy ezeket tényleg elfogadja. Másrészt, egy elfogadó számítás során amíg az A állapotban van, a bemeneti $w = a_1 a_2 \cdots a_n$ szó elejének megfelelően rak $\mathtt{a}$ és $\mathtt{b}$ betűket a verembe. 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: $S \to \mathtt{a}S\mathtt{a} \mid \mathtt{b}S\mathtt{b} \mid \varepsilon$

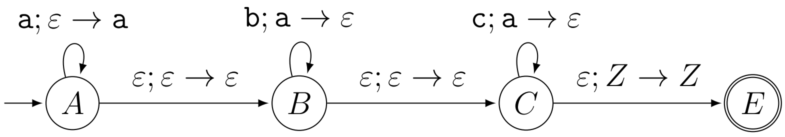
5. Adjon veremautomatát az alábbi nyelvekhez!
   (a) $L_a = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k : i, j, k \geq 0 \text{ és } i + j = k\}$
   (b) $L_b = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k : i, j, k \geq 0 \text{ és } j + k = i\}$
   (c) $L_c = \{\mathtt{a}^i\mathtt{b}^j\mathtt{c}^k : i, j, k \geq 0 \text{ és } i + k = j\}$

Megoldás: (a) Az ötlet az, hogy amikor $\mathtt{a}$ vagy $\mathtt{b}$ karakter jön, akkor mondjuk mindig egy $\mathtt{a}$-t rakunk a verembe, a $\mathtt{c}$-kre meg egyet kiszedünk. Akkor fogadunk el, ha a szó végén csak a Z marad a veremben. Persze arra is vigyáznunk kell, hogy $\mathtt{b}$ után nem jöhet $\mathtt{a}$, $\mathtt{c}$ után meg már csak $\mathtt{c}$ 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.

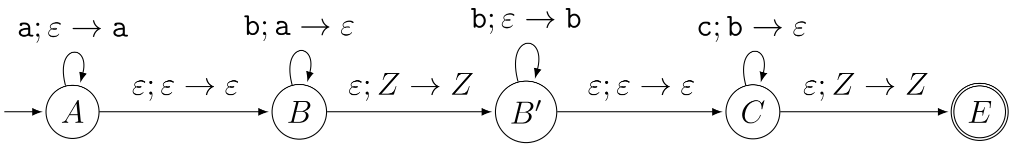


A különböző állapotok közötti $\varepsilon; \varepsilon \to \varepsilon$ átmenetek biztosítják azt is, hogy az üres szót ($i = j = k = 0$) elfogadja.

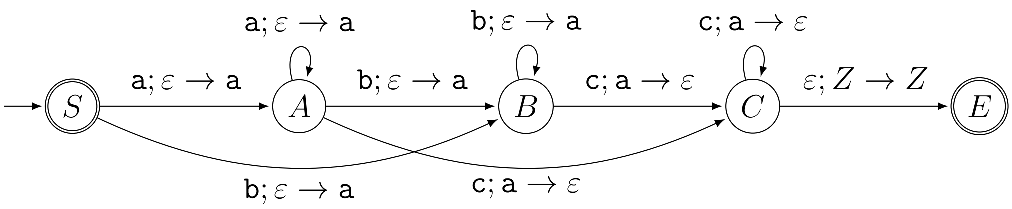
(b) Hasonló az előzőhöz, de most csak az a-kra rakunk be egy-egy a-t a verembe, b-re és c-re is kiveszünk egyet.



(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:



6. Adjon veremautomatát az alábbi nyelvekhez!
   (a) $L_a = \{\mathtt{a}^n\mathtt{b}^m : 2n = m \geq 1\}$ (b) $L_b = \{\mathtt{a}^n\mathtt{b}^m : 2n \geq m \geq n \geq 1\}$

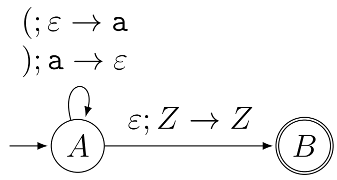
Megoldás: Itt most már csak szavakban vázoljuk a veremautomaták működését.
(a) A kezdő állapotból az első a olvasásakor két darab a betűt rak a verembe és átmegy az A állapotba. Itt minden további a betűnél megint két darab a betűt rak a verembe. Az A-ból az első b olvasásakor kivesz a veremből egy a-t 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.

7. 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.



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

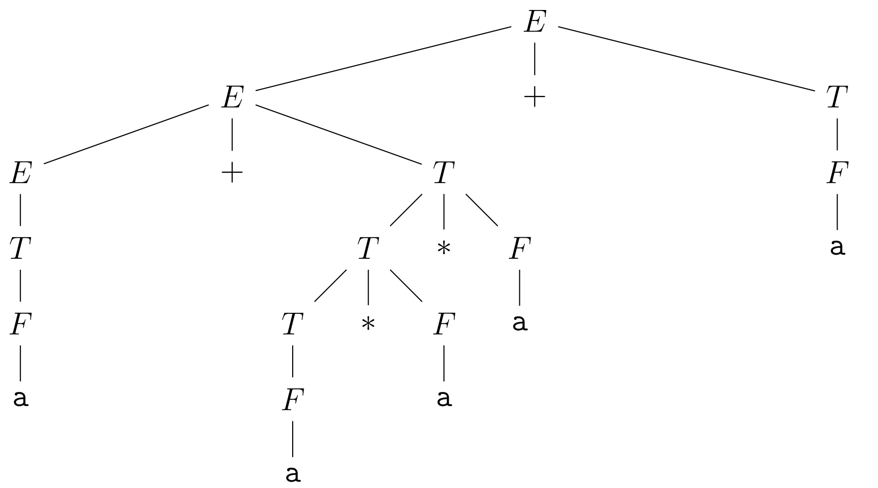
   (a) Adjon meg az $\mathtt{a}+\mathtt{a}*\mathtt{a}*\mathtt{a}+\mathtt{a}$ szóhoz egy levezetési fát!

   (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 $\mathtt{a}-\mathtt{a}+\mathtt{a}-\mathtt{a}$ levezetési fája? Mi a műveletek sorrendje?

Megoldás: (a)

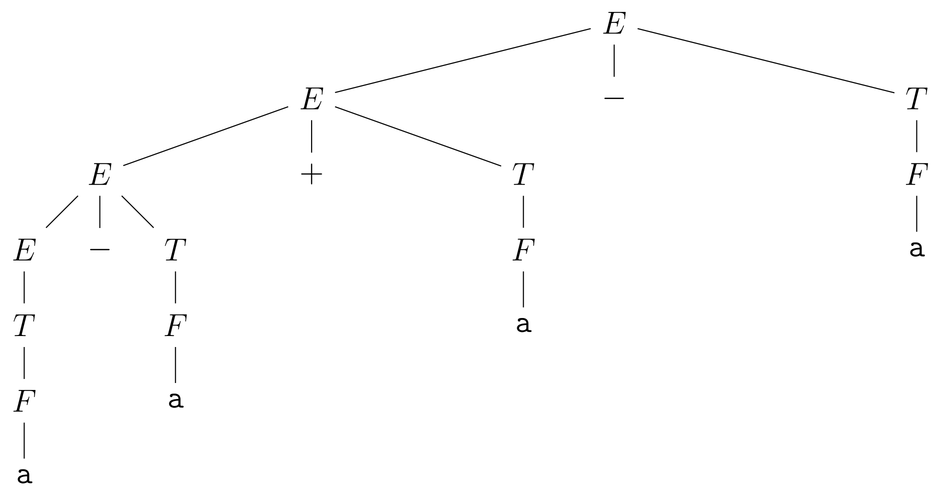


(b) Előbb a szorzások balról jobbra, utána az összeadások balról jobbra.

(c) A + és − egyenrangú műveletek, ezért ugyanarra a szintre kerülnek:

$$E \to E+T \mid E-T \mid T$$
$$T \to T*F \mid F$$
$$F \to (E) \mid \mathtt{a}$$

(d)



A kiértékelés a fa alapján balról jobbra történik.

9. Adjon meg egy környezetfüggetlen nyelvtant, ami az összes olyan helyes kifejezést generálja, amiben szorzás, hatványozás és zárójelek lehetnek! A nyelvtan legyen egyértelmű és tükrözze a műveletek szokásos sorrendjét (azaz, ha nincs zárójel, akkor előbb hatványozunk, és pl. a 2^{3^4}-nél előbb a 3^4-et kell kiszámolni és ez lesz a 2 kitevője).

   Megoldás: Két dologra kell figyelni: a hatványozás előbb történik, mint a szorzás és hogy több hatványozásnál a helyes sorrend a jobbról balra haladás. Ezért az első szabály a szorzás, második a hatványozás, és a hatványozásnál az eddigiekhez képest a két változó sorrendje fordított lesz, azaz

$$
\begin{aligned}
T &\to T * F \mid F \\
F &\to B \uparrow F \mid B \\
B &\to (T) \mid \mathtt{a}
\end{aligned}
$$