

1. Mit generál az $X \rightarrow cXd \mid Y; \quad Y \rightarrow cY \mid c$ nyelvtan?

Megoldás: Sejtés: ez a nyelvtan az olyan $c^i d^j$ alakú szavakat generálja, amikben $j \geq 0$ és $i > j$.

Ennek belátására két dolgot kell megmutatni:

Csak ilyen szavakat tud generálni:

A levezetés elején egy darabig az X baloldali első szabályt használjuk j -szer ($j \geq 0$), eközben $c^j X d^j$ alakú szavak állnak elő. Ezután, ha terminális karakterekből álló szót akarunk, akkor használunk kell az $X \rightarrow Y$ szabályt, így kapunk egy $c^j Y d^j$ alakú szót. Ezután y -ből tetszőleges számú, de legalább egy darab c betűt generálunk, így végül egy $c^j c^t d^j$ alakú szót kapunk, ahol $j \geq 0$ és $t \geq 1$, azaz pont egy olyan szó, amit sejtettünk.

Minden ilyen szót generál:

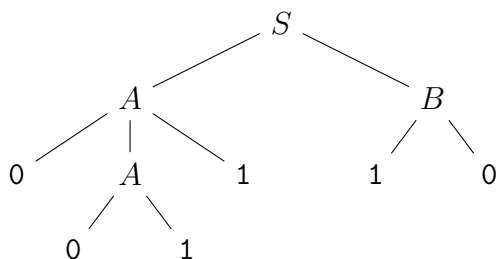
Ha kapunk egy olyan $c^i d^j$ alakú szót, amiben $j \geq 0$ és $i > j$, akkor az előző részben vázolt levezetéssel ezt elő is tudjuk állítani.

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

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

(b) Határozza meg a nyelvtan által generált nyelvet!

Megoldás: (a)



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

3. Határozza meg az $S \rightarrow A \mid B; \quad A \rightarrow 0A1 \mid 01; \quad B \rightarrow 1B0 \mid 10$ nyelvtan által generált nyelvet!

Megoldás: a nyelv az A -ból, illetve a B -ből generálható nyelvek uniója, és ezért

$L = \{0^n 1^n \mid n \geq 1\} \cup \{1^n 0^n \mid n \geq 1\}$.

4. Adjon környezetfüggetlen nyelvtant a következő nyelvekre!

(a) $(0 + 1)^* 011(0 + 1)^*$

(b) $1(0 + 1)^* 0$

(c) $((0 + 1)(0 + 1))^*$

Megoldás: Sok helyes nyelvtan van, mutatunk egyet-egyet, ami a reguláris kifejezés szerkezetét tükrözi.

(a) $S \rightarrow A011A; \quad A \rightarrow 0A \mid 1A \mid \varepsilon$ (A -ból a $(0 + 1)^*$ generálható, előadáson is volt)

(b) $S \rightarrow 1A0; \quad A \rightarrow 0A \mid 1A \mid \varepsilon$

(c) $S \rightarrow 00S \mid 01S \mid 10S \mid 11S \mid \varepsilon$, de például $S \rightarrow 0S0 \mid 0S1 \mid 1S0 \mid 1S1 \mid \varepsilon$ is jó.

5. Adjon környezetfüggetlen nyelvtant az alábbi nyelvekre!

(a) $\{a^n b^{n+1} \mid n \geq 0\}$

(b) $\{a^n b^{2n} \mid n \geq 0\}$

(c) $\{a^n b^m \mid 0 \leq n \leq m \leq 2n\}$

Megoldás: Mindegyiknél több jó megoldás is van, pl.:

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

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 $S \rightarrow a S b \mid b$.

(b) Az ötlet, hogy most egy lépésben egy a betűt teszünk előre és két b -t hátra: $S \rightarrow a S b b \mid \varepsilon$.

Világos, hogy az így kapott nyelvtan jó.

(c) Az ötlet, hogy most egy lépésben egy a betűt teszünk előre és egy vagy két b -t hátra:

$S \rightarrow a S b \mid a S b b \mid \varepsilon$.

Indukcióval könnyen látható, hogy az a és b betűk száma közötti egyenlőtlenségek végig fennállnak.

6. Adjon környezetfüggetlen nyelvtant a jó zárójelezések nyelvéhez!

Megoldás: A zárójelsorozatot elképzelhetjük úgy, hogy vannak a külső szintű zárójelek (ezekből akárhány), és minden ilyen külső szintű zárójelpáron belül is egy jó sorozatnak kell lenni. Ebből a következő nyelvtant kaphatjuk:

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

Az első szabállyal legyárthatjuk a tetszőleges számú külső szintű zárójel helyét, a második szabály kirakja a zárójelpárokat, és lehetőséget ad, hogy a belsejében folytassuk az eljárást.

Kezdhethetjük az első külső zárójelpárral is, aminek a belsejében, és utána is jó sorozatnak kell lenni:

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

Vagy kezdhethetjük egy tetszőleges külső zárójelpárral. Ekkor előtte, benne és utána is helyes sorozat kell álljon:

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

7. Határozza meg az alábbi környezetfüggetlen nyelvtanok által generált nyelveket!

$$(a) \quad T \rightarrow TT \mid aTb \mid bTa \mid a \mid \varepsilon$$

$$(b) \quad R \rightarrow TaT; \quad T \rightarrow TT \mid aTb \mid bTa \mid a \mid \varepsilon$$

Megoldás: (a) Az világos, hogy a keletkezett szóban, ha nem az üres szó, akkor legalább annyi **a** betű lesz, mint **b**. (Ez utóbbi tulajdonság valójában az üres szóra is teljesül.) Megmutatjuk, hogy minden ilyen szó levezethető, ezt a hossz szerinti teljes indukcióval csináljuk. Tekintsünk egy ilyen w szót. Ha a hossza $|w| \leq 1$, akkor vagy $w = \varepsilon$ vagy $w = a$, és mindkettő valóban levezethető. Tegyük fel, hogy minden n -nél rövidebb, legalább annyi **a** betűt, mint **b** betűt tartalmazó szó levezethető.

Legyen most $w = x_1x_2 \cdots x_n$ egy $n \geq 2$ hosszú szó, amiben legalább annyi **a** van, mint **b**.

Legyen i a legkisebb olyan pozitív szám, melyre teljesül, hogy az $x_1 \cdots x_i$ részsóban ugyanannyi **a** van, mint **b**.

Ha nincs ilyen i , akkor minden kezdőszeletben, és az egész szóban is több az **a**, mint a **b**. Tehát biztos, hogy $x_1 = a$ és ha ezt a betűt levágjuk, akkor is a nyelvben maradunk, ezért az indukciós feltevés miatt a $T \Rightarrow TT \Rightarrow aT$ kezdés folytatható úgy, hogy a végén jó levezetést kapjunk.

Vegyük észre, hogy ha van jó i , akkor $x_1 \neq x_i$ (mert az i -ediknél lesz pont ugyanannyi **a**, mint **b**). Ezért ha a levezetés első lépése után az első T -ből a 2. vagy 3. szabállyal megkapjuk az x_1 és x_i karaktereket, közéjük a többi (itt is ugyanannyi **a** van, mint **b**) a T -ből generálható. A szó végén is legalább annyi **a** van, mint **b**, ezért ez megkapható az első lépésben kapott második T -ből.

Másik megfontolás („alulról felfelé”): kiindulunk a szóból, és alkalmazzuk a következő átírási szabályokat: tetszőleges ab vagy ba részsót helyettesítsünk T -vel (a $T \Rightarrow aTb \Rightarrow ab$ vagy a $T \Rightarrow bTa \Rightarrow ba$ lépéssorozatok megfordításai); TT -t helyettesítsünk T -vel; aTb és bTa szintén helyettesíthető T -vel. Mit kapunk, amikor ezek egyike sem alkalmazható: **b** betű biztos nem marad (mert akkor **a** is kell, hogy maradjon, és lesz, esetleg T -vel elválasztott, **a** és **b** is). Ha már csak **a** és T maradt, helyettesítsük az **a** betűket T -vel, a szomszédos T -ket meg egyetlen T -vel. Így a végén egyetlen T marad csak, és ekkor megkaptunk (visszafele) egy levezetést.

(b) A nyelv azokból a szavakból áll, amelyekben több **a** van, mint **b**.

Vegyük észre, hogy a T -re az előző szabályok maradtak, azaz T -ből azok a szavak vezethetők le, amelyekben legalább annyi **a** van, mint **b**. A kezdő szabállyal még egy **a** betűt hozzáteszünk, tehát biztos, hogy a levezetett szavakban több **a** lesz, mint **b**. Azt kell még megindokolni, hogy minden ilyen megkapunk: Egy ilyen w szót bontsunk úgy fel, hogy $w = w_1aw_2$, ahol w_1 -ben ugyanannyi **a** van, mint **b**, w_2 -ben meg legalább ugyanannyi. Ez a w_1 és w_2 is levezethető T -ből, tehát az egész szó is.

Ilyen felbontás mindig van, mert tekintsük a legrövidebb kezdőszeletet, amiben az **a**-k száma több, mint a **b**-k száma. Ennek az utolsó karaktere **a**, az ez előtti rész legyen w_1 , az utána levő pedig w_2 . (Lehet, hogy $w_1 = \varepsilon$ vagy $w_2 = \varepsilon$.)

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

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

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

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: $S \rightarrow aSa \mid bSb \mid a \mid b \mid \varepsilon$

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 -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.)

9. 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 -ból a $0^k \# 0^{2k}$ alakúak ($k \geq 0$). Tehát a nyelv: $L = \{0^k \# 0^t \# 0^s \mid k, t, s \geq 0\} \cup \{0^k \# 0^{2k} \mid k \geq 0\}$. 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 többi a másodikból, de úgy is, hogy a $0\#$ részt vezetjük csak le az első T -ből és a többi 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 ilyet csinálni a nyelv ismeretében: az U jó, ahogy van, a T változót cseréljük le például így $S \rightarrow N\#N\#N \mid U, N \rightarrow 0N \mid \varepsilon, U \rightarrow 0U00 \mid \#$

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

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

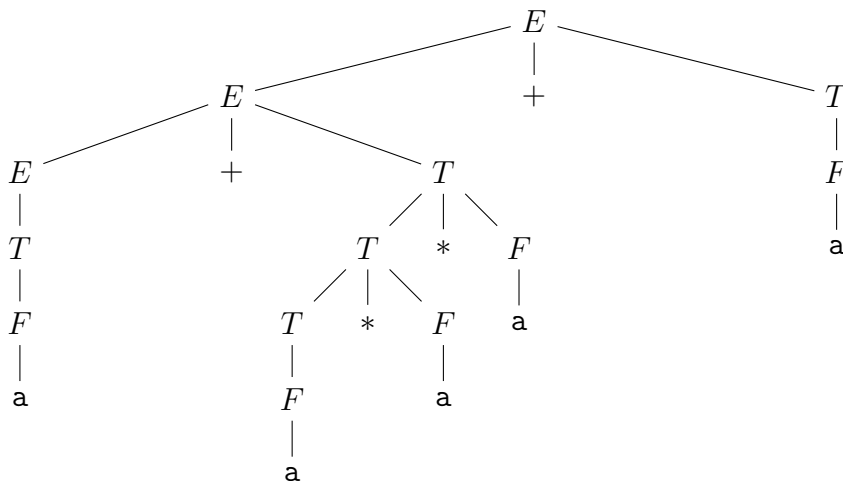
(a) Adjon meg az $a + a * a * a + a$ szóhoz egy levezetési fát!

(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?

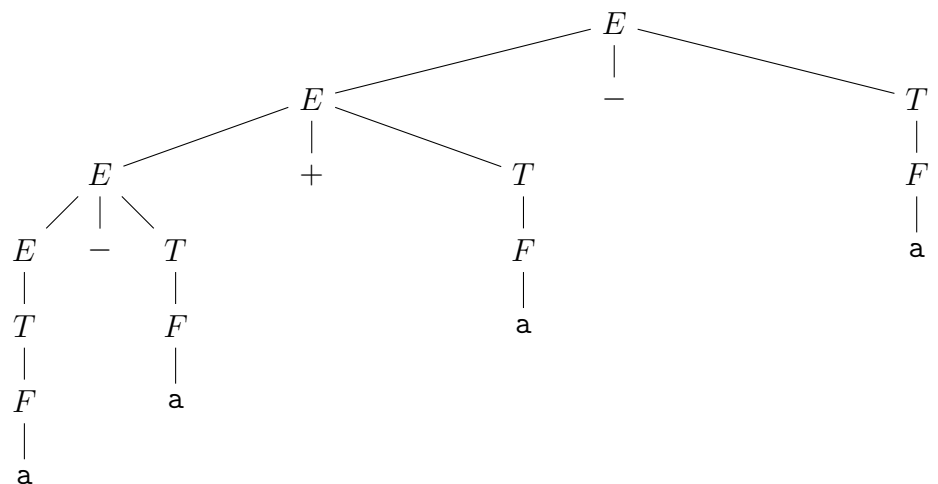
Megoldás: (a)



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

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

$$\begin{aligned} E &\rightarrow E + T \mid E - T \mid T \\ T &\rightarrow T * F \mid F \\ F &\rightarrow (E) \mid a \end{aligned}$$



A kiértékelés a fa alapján balról jobbra történik.

11. 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

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