

Sajátos nyelvtanok és nyelvek

- $LL(k)$,
- $LR(k)$,
- elsőbbségi (precedencia) nyelvtanok

$LL(k)$ nyelvtanok

Ahhoz, hogy bebizonyítsuk, hogy egy szó benne van egy környezetfüggetlen nyelvben, meg kell adnunk az adott szónak egy levezetését (derivációját). Ez általában nem könnyű feladat, hisz egy-egy alkalommal több helyettesítés közül is választhatunk, és nem könnyű eldönteni, hogy melyiket válasszuk. Ha ez a választás mindig megtörténhet egyértelműen úgy, hogy az elemzendő szóban a még meg nem vizsgált rész k ($k \geq 1$) betűjét előre megnézzük, akkor azt mondjuk, hogy az illető nyelvtan $LL(k)$ típusú, az elemzés pedig $LL(k)$ elemzés.

Az $LL(0)$ eset nem érdekes, mert ez a nyelvtan csupán egyetlen szót generál.

Az $LL(k)$ rövidítés a következő szövegből származik: **Left to right scan, producing Leftmost derivation with k symbol lookahead.**

Mielőtt még definiálnánk az $LL(k)$ nyelvtant, vezessük be a következő jelölést egy adott $G = (N, T, P, S)$ nyelvtanra és egy $\alpha \in (NUT)^*$ szóra.

$$FIRST_k^G(\alpha) = \{x \mid \alpha \xrightarrow[G]{*} xy, xy \in T^*, |x| = k\} \cup \{x \mid \alpha \xrightarrow[G]{*} x, x \in T^*, |x| < k\}.$$

azaz, $FIRST_k^G(\alpha)$ az α -ból G -ben levezethető terminális szavak k hosszúságú kezdőszeleteinek a halmazát jelenti. Ha a terminális szó rövidebb k -nál, akkor azonos a k hosszúságú kezdőszeletével.

Ha $\alpha \in T^*$, akkor $\alpha = xy$ vagy $\alpha = x$, tehát $FIRST_k^G(\alpha)$ az α terminális szó k hosszúságú kezdőszeletét tartalmazza csupán.

A G felső indexet esetleg elhagyhatjuk.

Példa.

$$\begin{aligned}
 G &= (\{S\}, \{a, b\}, \{S \rightarrow ab, S \rightarrow aSb\}, S) \\
 FIRST_1(aSb) &= \{a\} \\
 FIRST_2(aSb) &= \{aa\} \\
 FIRST_3(aSb) &= \{aaa, aab\} \\
 FIRST_4(aSb) &= \{aabb, aaab, aaaa\} \\
 FIRST_5(aSb) &= \{aabb, aaabb, aaaab, aaaaa\}
 \end{aligned}$$

A $G = (N, T, P, S)$ környezetfüggetlen nyelvtant $LL(k)$ nyelvtannak ($k \geq 0$) nevezzük, ha bármely

$$S \xrightarrow[G]{*} uA\alpha \xrightarrow[G]{*} u\beta_1\alpha \xrightarrow[G]{*} ux$$

$$S \xrightarrow[G]{*} uA\alpha \xrightarrow[G]{*} u\beta_2\alpha \xrightarrow[G]{*} uy$$

legbaloldalibb levezetéspárra (ahol $\alpha, \beta_1, \beta_2 \in (N \cup T)^*$, $u, x, y \in T^*$) a $FIRST_k(x) = FIRST_k(y)$ egyenlőségből következik, hogy $\beta_1 = \beta_2$.

Példa. Vizsgáljuk meg a $G = (\{S\}, \{a, b\}, \{S \rightarrow ab, S \rightarrow aSb\}, S)$ nyelvtant!

A következő levezetéspárból következik, hogy G nem $LL(1)$ nyelvtan.

$$S \Longrightarrow aSb \xRightarrow{*} \underbrace{aa}_u \underbrace{S}_A \underbrace{bb}_\alpha \Longrightarrow \underbrace{aa}_u \underbrace{ab}_{\beta_1} \underbrace{bb}_\alpha \xRightarrow{*} \underbrace{aa}_u \underbrace{abbb}_x$$

$$S \Longrightarrow aSb \xRightarrow{*} \underbrace{aa}_u \underbrace{S}_A \underbrace{bb}_\alpha \Longrightarrow \underbrace{aa}_u \underbrace{aSb}_{\beta_2} \underbrace{bb}_\alpha \xRightarrow{*} \underbrace{aa}_u \underbrace{aabbbb}_y$$

És annak ellenére, hogy $FIRST_1(x) = FIRST_1(y) = \{a\}$, nem következik a $\beta_1 = \beta_2$ egyenlőség, mivel $\beta_1 = ab$, $\beta_2 = aSb$.

Tétel. Egy $G = (N, T, P, S)$ környezetfüggetlen nyelvtan akkor és csak akkor $LL(k)$ típusú ($k \geq 1$), ha bármely $S \xRightarrow{*} uA\gamma$ legbaloldali levezetésre (ahol $u \in T^*$, $\gamma \in (N \cup T)^*$) és $A \rightarrow \alpha$, $A \rightarrow \beta$ két különböző szabályra

$$FIRST_k(\alpha\gamma) \cap FIRST_k(\beta\gamma) = \emptyset.$$

Példa. Előbbi példánk, $G = (\{S\}, \{a, b\}, \{S \rightarrow ab, S \rightarrow aSb\}, S)$ esetében a következő levezetések lehetségesek:

$$S \xRightarrow{*} a^n S b^n, \quad n \geq 1 \text{ és}$$

$$S \xRightarrow{*} ab$$

valamint léteznek az $S \rightarrow aSb$ és $S \rightarrow ab$ szabályok.

Ekkor

$$FIRST_1(aSb^n) = \{a\}, \quad FIRST_1(ab^n) = \{a\}; \text{ metszetük nem üres.}$$

Tehát G nem $LL(1)$ típusú.

$FIRST_2(aSbb^n) = \{aa\}$, $FIRST_2(abb^n) = \{ab\}$; metszetük üres.
Tehát a G nyelvtan $LL(2)$ típusú.

Könnyű belátni, éppen a fenti tétel alapján, hogy ha egy nyelvtan $LL(k)$, akkor egyben $LL(k+1)$ is. Például az előbbi nyelvtan esetében, amely $LL(2)$ típusú:

$FIRST_3(aSbb^n) = \{aaa, aab\}$, $FIRST_3(abb^n) = \{aba\}$, metszetük üres.
Tehát a G nyelvtan $LL(3)$ típusú is.

Következmény. Ha egy környezetfüggetlen nyelv $LL(k)$ típusú, akkor $LL(k+1)$ típusú is.

Fontos eset az $LL(1)$. Ekkor annak vizsgálata, hogy egy nyelvtan $LL(1)$ típusú nagyon egyszerű, mivel nyelvtanaink ε -mentesek. (Nincs $A \rightarrow \varepsilon$ szabály, kivéve ha $S \rightarrow \varepsilon$, de ekkor S szabály jobb oldalán nem szerepelhet.)

Tétel. Egy környezetfüggetlen nyelvtan akkor és csakis akkor $LL(1)$ típusú, ha minden $A \rightarrow \alpha$, $A \rightarrow \beta$ különböző szabálypárra
 $FIRST_1(\alpha) \cap FIRST_1(\beta) = \emptyset$.

Példa. 1) A $G_1 = \{S, A, B\}, \{a, b\}, P, S)$, ahol

$$P : \begin{aligned} S &\rightarrow aAB \mid bBA \\ A &\rightarrow a \mid bS \\ B &\rightarrow b \mid aS \end{aligned}$$

$LL(1)$ típusú nyelvtan, mivel:

$$FIRST_1(aAB) = \{a\}, \quad FIRST_1(bBA) = \{b\}$$

$$FIRST_1(a) = \{a\}, \quad FIRST_1(bS) = \{b\}$$

$$FIRST_1(b) = \{b\}, \quad FIRST_1(aS) = \{a\} \text{ és a megfelelő metszetek}$$

üresek.

2) a $G_2 = (\{S, A\}, \{0, 1\}, \{S \rightarrow 0A \mid 0, A \rightarrow 0A \mid 0A1 \mid 01 \mid 1\}, S)$ nyelvtan nem $LL(1)$, mivel például az $A \rightarrow 0A, A \rightarrow 0A1$ szabályokra:

$FIRST_1(0A) = \{0\}, FIRST_1(0A1) = \{0\}$, tehát metszetük nem üres.

A következő tétel fontos az elemzés szempontjából, és az értelmezésből következik.

Tétel. Minden $LL(k)$ nyelvtan egyértelmű.

Egy A változót (nemterminális jelet) **balrekurzív** nevezünk, ha létezik

$$A \xrightarrow{*} A\alpha \quad (\alpha \in (N \cup T)^*)$$

levezetés.

Ha egy környezetfüggetlen nyelvtanban van olyan változó, amely nem szerepel egyetlen terminális szó levezetésében sem, akkor az a nyelvtan **redundáns**.

Tétel. Ha egy G nem redundáns környezetfüggetlen nyelvtanban van balrekurzív változó, akkor G semmilyen k -ra sem $LL(k)$.

Példa. Legyen adva a következő nyelvtan:

$$G = (\{S\}, \{a, b\}, \{S \rightarrow Sa, S \rightarrow b\}, S).$$

Ekkor $S \xrightarrow{*} Sa^n$. Meg kell vizsgálnunk a következő halmazokat:

$$FIRST_k(Saa^n) = \{ba^{k-1}\} \text{ és } FIRST_k(ba^n) = \{ba^{k-1}\}.$$

Látszik, hogy egybeesnek tetszőleges $k \geq 1$ értékre, tehát a metszetük nem lehet üres. A nyelvtan nem $LL(0)$ sem, így G nem $LL(k)$ egyetlen k -ra sem.

A $FIRST_k(\alpha)$ halmazok meghatározása

Először meghatározzuk a $FIRST_k(A)$ halmazokat, ahol A változó.

Ehhez generáljuk az összes $A \xrightarrow{*} u\alpha$ alakú levezetéseket, ahol $u \in T^*$ és $|u| \geq k$.

Ha $a \in T$, akkor $FIRST_k(a) = \{a\}$.

Tetszőleges $\alpha = X_1X_2 \dots X_n$ és $X_i \in N \cup T$ ($i = 1, 2, \dots, n$) esetében:

$$FIRST_k(X_1X_2 \dots X_n) = FIRST_k\left(FIRST_k(X_1)FIRST_k(X_2) \dots FIRST_k(X_n)\right),$$

ahol a jobb oldali zárójelben szóhalmazok szorzata (konkatenációja) szerepel.

A $FIRST_1(X)$, $X \in N$ meghatározására létezik egy jól algoritmizálható módszer, amely relációk tranzitív lezártjának a kiszámításán alapszik.

Értelmezzük a $G = (N, T, P, S)$ nyelvtan esetében a következő bináris relációt:

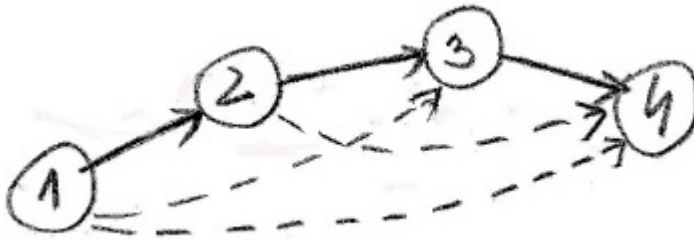
$$\rho \subseteq N \times (N \cup T), \quad X\rho Y, \quad \text{ha létezik az } X \rightarrow Y\alpha \text{ szabály.}$$

Ennek a relációnak a mátrixára alkalmazzuk a Warshall-algoritmust a tranzitív lezárt kiszámítására. A mátrix $A = (a_{ij})_{i=\overline{1,m}, j=\overline{1,n}}$, ahol

$$a_{ij} = \begin{cases} 1 & \text{ha } X_i\rho Y_j \text{ (} X_i \in N, Y_j \in N \cup T\text{)} \\ 0 & \text{különben} \end{cases}$$

A Warshall-algoritmus esetünkben a következőképpen írható le. Az algoritmus bemenete a reláció A mátrixa, ahol $j = 1, 2, \dots, m$ -re $Y_j = X_j$ ($|N| = m$, $|N \cup T| = n$), eredménye pedig a tranzitív lezárt R mátrixa.

tranzitív lezárt:



WARSHALL(A)

1. $R := A$
2. **for** $k = 1$ **to** m
3. **do for** $i = 1$ **to** m
4. **do for** $j = 1$ **to** n
5. **do if** $r_{ik} = 1$ **és** $r_{kj} = 1$
6. **then** $r_{ij} := 1$
7. **return** R

Ekkor

$$FIRST_1(X_i) = \{Y_j \mid r_{ij} = 1, j = m + 1, \dots, n\}, \quad i = 1, 2, \dots, m$$

Példa.

Legyen $G = (\{E, T, F\}, \{a, +, *, (,)\}, P, E)$, ahol a szabályok a következők:

$$E \rightarrow E + T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow (E) \mid a$$

Ekkor az A mátrix a következő:

	E	T	F	$+$	$*$	$($	$)$	a
E	1	1	0	0	0	0	0	0
T	0	1	1	0	0	0	0	0
F	0	0	0	0	0	1	0	1

Az eredménymátrix pedig, amelyet a Warshall-algoritmussal kapunk meg:

	E	T	F	$+$	$*$	$($	$)$	a
E	1	1	1	0	0	1	0	1
T	0	1	1	0	0	1	0	1
F	0	0	0	0	0	1	0	1

Innen pedig azt kapjuk, hogy:

$$FIRST_1(E) = \{ (, a \}, \quad FIRST_1(T) = \{ (, a \} \quad FIRST_1(F) = \{ (, a \}$$

Ez a nyelvtan nem $LL(1)$ típusú, hisz például az $E \rightarrow E + T$ és $E \rightarrow T$ szabályokra

$$FIRST_1(E + T) = FIRST_1(T) = \{ (, a \}.$$

Bizonyítás nélkül közöljük a következő eredményeket:

- 1) Tetszőleges k -ra létezik olyan algoritmus, amely eldönti egy nyelvtanról, hogy $LL(k)$ típusú-e.
- 2) Tetszőleges pozitív k -ra létezik olyan $LL(k+1)$ nyelvtan, amelyik nem ekvivalens egyetlen $LL(k)$ nyelvtannal sem.

$LR(k)$ nyelvtanok

Hasonlóan az $LL(k)$ nyelvtanokhoz, értelmezhetjük az $LR(k)$ nyelvtanokat, de itt legbaloldalibb levezetések helyett legjobboldalibbakat használunk. Ezért van nevében az L (**leftmost**) helyett R (**rightmost**).

A legjobboldalibb levezetésekénél mindig a szóban lévő utolsó (legjobboldalibb) változót helyettesítjük.

Az $LR(k)$ nyelvtanok esetében a szavak felismerését „fordított” levezetéssel (redukálással) valósítjuk meg. Azaz a szóban megkeressük egy szabály jobb oldalát, majd helyettesítjük a bal oldali változóval, azaz redukáljuk a bal oldali változóra. Ezt mindaddig végezzük, amíg el nem érünk a nyelvtan kezdőszimbólumához. Vegyünk például a

$$G = (\{A, B\}, \{a, +, *\}, \{A \rightarrow A + B \mid B, B \rightarrow B * B \mid a\}, A)$$

nyelvtant. Vizsgáljuk meg, hogy az $a + a * a$ szó eleme-e $L(G)$ -nek. A következő redukálásokat végezzük:

$$\begin{array}{ll} a + a * a & \\ B + a * a, & \text{mivel létezik } B \rightarrow a \text{ szabály,} \\ A + a * a, & \text{mivel létezik } A \rightarrow B \text{ szabály,} \\ A + B * a, & \text{mivel létezik } B \rightarrow a \text{ szabály,} \\ A + B * B, & \text{mivel létezik } B \rightarrow a \text{ szabály,} \\ A + B, & \text{mivel létezik } B \rightarrow B * B \text{ szabály,} \\ A, & \text{mivel létezik } A \rightarrow A + B \text{ szabály.} \end{array}$$

Mivel megkaptuk a nyelvtan kezdőszimbólumát, a szó benne van a G által generált nyelvben. Természetesen, nem mindig könnyű eldönteni, hogy hol melyik jobb oldalt helyettesítsük, ha több lehetőség is van.

Ha például, a negyedik sorban $A + B$ -t helyettesítjük A -val, nem jutunk helyes eredményre.

Az $LR(k)$ nyelvtanok pont ebben segítenek, mégpedig úgy, hogy k betű előre megnézésével el lehet dönteni, hogy melyik szabályt kell alkalmaznunk. A fenti G nyelvtan $LR(1)$ típusú. Például a negyedik sorban, ha a B betű után megnézzük a következőt (itt $*$), eldönthetjük, hogy tovább kell lépniük, ha $+$ jel lett volna, akkor lehetett volna alkalmazni redukálásra az $A \rightarrow A + B$ szabályt.

A fenti redukálássorozat („fordított” levezetés) megfelel egy legjobboldali levezetésnek.

$$\begin{aligned} A &\Longrightarrow A + B \Longrightarrow A + B * B \Longrightarrow A + B * a \Longrightarrow A + a * a \Longrightarrow \\ &\Longrightarrow B + a * a \Longrightarrow a + a * a. \end{aligned}$$

Csak olyan nyelvtanokkal foglalkozunk, amelyekben egyetlen szabály kezdődik a kezdőszimbólummal, és ez nem szerepel egyetlen szabály jobb oldalán se.

Egy $G = (N, T, P, S)$ környezetfüggetlen nyelvtant $LR(k)$ nyelvtannak ($k \geq 0$) nevezünk, ha tetszőleges

$$\begin{aligned} S &\xrightarrow{*} \alpha Ax \Longrightarrow \alpha \beta x \\ S &\xrightarrow{*} \gamma By \Longrightarrow \gamma \delta y = \alpha \beta z \end{aligned}$$

legjobboldalibb levezetéspárra abból, hogy $FIRST_k(x) = FIRST_k(z)$ következik, hogy $\alpha = \gamma$, $\beta = \delta$, $A = B$ (és következésképpen $y = z$) (ahol $\alpha, \beta, \gamma \in (N \cup T)^*$, $x, y, z \in T^*$, $A, B \in N$).

Példák.

1) A $G_1 = (\{S, A\}, \{a\}, \{S \rightarrow A, A \rightarrow aAb \mid aa \mid a\}, S)$ nyelvtan nem $LR(k)$ egyetlen k -ra sem, mert

$$S \Rightarrow A \xRightarrow{*} \underbrace{a^k}_{\alpha} \underbrace{A}_A \underbrace{b^k}_x \Rightarrow a^{k+2}b^k = \underbrace{a^k}_{\alpha} \underbrace{aa}_{\beta} \underbrace{b^k}_x$$

$$S \Rightarrow A \xRightarrow{*} \underbrace{a^{k+1}}_{\gamma} \underbrace{A}_B \underbrace{b^{k+1}}_y \Rightarrow a^{k+2}b^{k+1} = \underbrace{a^{k+1}}_{\gamma} \underbrace{a}_{\delta} \underbrace{b^{k+1}}_y = \underbrace{a^k}_{\alpha} \underbrace{aa}_{\beta} \underbrace{b^{k+1}}_z$$

és $FIRST_k(x) = FIRST_k(z)$, de ennek ellenére $\alpha = a^k \neq a^{k+1} = \gamma$, tehát G_1 semmilyen k -ra sem lehet $LR(k)$.

2) $G_2 = (\{S, E, T\}, \{a, +, *\}, \{S \rightarrow E, E \rightarrow E+T \mid T, T \rightarrow T*T \mid a\}, S)$
Ez a nyelvtan nem $LR(0)$, mert

$$S \xRightarrow{*} \underbrace{\varepsilon}_{\alpha} \underbrace{S}_A \underbrace{\varepsilon}_x \Rightarrow \underbrace{\varepsilon}_{\alpha} \underbrace{E}_{\beta} \underbrace{\varepsilon}_x$$

$$S \xRightarrow{*} \underbrace{\varepsilon}_{\gamma} \underbrace{E}_B \underbrace{\varepsilon}_y \Rightarrow \underbrace{\varepsilon}_{\gamma} \underbrace{E+T}_{\delta} \underbrace{\varepsilon}_y = \underbrace{\varepsilon}_{\alpha} \underbrace{E}_{\beta} \underbrace{+T}_z$$

és $FIRST_0(x) = FIRST_0(z) = \emptyset$, ennek ellenére $y \neq z$.

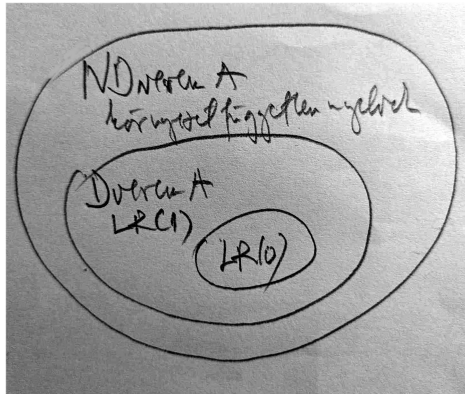
A G_2 nyelvtan ellenben $LR(1)$. Egy betű előreolvasásával mindig el lehet dönteni, hogy melyik szabályt kell alkalmazni.

Egy nyelv $LR(k)$ (vagy $LL(k)$) típusú, ha létezik olyan $LR(k)$ (vagy $LL(k)$) nyelvtan, amelyik generálja.

Bizonyítás nélkül közöljük a következő eredményeket.

- 1) Létezik olyan algoritmus, amely tetszőleges k -ra egy nyelvtanról eldönti, hogy $LR(k)$ típusú-e.
- 2) Minden $LR(k)$ nyelvtanhoz megadható egy vele ekvivalens $LR(1)$ nyelvtan.
- 3) Minden $LL(k)$ nyelvtan egyben $LR(k)$ is, de fordítva ez nem igaz. Például egy $LL(0)$ nyelvtan csupán egyetlen szót generál, míg egy $LR(0)$ akár végtelen nyelvet is generálhat.
- 4) Az $LR(1)$ nyelvek osztálya megegyezik a **determinisztikus veremautomaták** által felismert nyelvek (ún. determinisztikus környezetfüggetlen nyelvek) osztályával.
- 5) Az $LR(0)$ nyelvek osztálya megegyezik a **kezdőszeletmentes** determinisztikus környezetfüggetlen nyelvek osztályával.

Egy L nyelv akkor kezdőszeletmentes, ha bármely $u \in L$ szóra igaz az, hogy u egyetlen valódi kezdőszelete (prefixe) sem eleme L -nek.



nondeterminisztikus veremautomaták által felismert nyelvek
(**környezetfüggetlen nyelvek**)

↓

determinisztikus veremautomaták által felismert nyelvek
(determinisztikus környezetfüggetlen nyelvek)
(**LR(1) nyelvek**)

↓

kezdőszeletmentes determinisztikus környezetfüggetlen nyelvek
(**LR(0) nyelvek**)

Elsőbbségi (precedencia) nyelvtanok

Az ábécé betűi között relációkat határozunk meg, és ezek segítségével lehet eldönteni, hogy milyen redukálásokat kell a vizsgálandó szóban végezni egy alulról felfelé történő elemzésben.

Az egyszerűség kedvéért bevezetünk még egy új határoló jelet, a $\$$ -t, amely közrefogja a vizsgálandó szót.

A következő relációkat használjuk \prec (kisebb), \succ (nagyobb) és \doteq (egyenlő). Ezek nem feltétlenül tranzitív, illetve ekvivalencia relációk, és két jel között többféle reláció is létezhet. A relációkat a következőképpen értelmezzük.

- $\$ \prec X$ ha létezik $S \xRightarrow{+} X\beta$ alakú levezetés,
- $X \prec Y$ ha létezik $A \rightarrow \alpha X B \beta$ alakú szabály, és $B \xRightarrow{+} Y\gamma$,
- $X \succ a$ ha létezik $A \rightarrow \alpha B Y \beta$ alakú szabály, és $B \xRightarrow{+} \gamma X$, $Y \xRightarrow{*} a\delta$
- $X \doteq Y$ ha létezik $A \rightarrow \alpha X Y \beta$ alakú szabály,
- $X \succ \$$ ha létezik $S \xRightarrow{+} \alpha X$ alakú levezetés,

ahol $X, Y \in N \cup T$, $a \in T$, $A, B \in N$, $\alpha, \beta, \gamma, \delta \in (N \cup T)^*$.

A \prec , \doteq és \succ relációk meghatározására vezessük be a következő relációkat, ahol $A \in N$, $X \in N \cup T$, $\alpha, \beta \in (N \cup T)^*$:

- $A \varrho_l X$ ha létezik $A \rightarrow X\beta$ szabály,
- $A \varrho_r X$ ha létezik $A \rightarrow \alpha X$ szabály.

Ezen relációkat az $L = (l_{ij})$ és $R = (r_{ij})$ mátrixok segítségével ábrázoljuk.

Használjuk a következő jelöléseket:

$$N = \{X_1, X_2, \dots, X_m\},$$

$$N \cup T = \{Y_1, Y_2, \dots, Y_n\} \quad (\text{ahol } Y_i = X_i, \text{ ha } i = 1, 2, \dots, m).$$

$$l_{ij} = \begin{cases} 1, & \text{ha } X_i \varrho_l Y_j, \\ 0, & \text{különben.} \end{cases} \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n$$

$$r_{ij} = \begin{cases} 1, & \text{ha } X_i \varrho_r Y_j, \\ 0, & \text{különben.} \end{cases} \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n$$

A továbbiakban a számításokat következő példán mutatjuk be.

Legyen $G = (\{S, A, B\}, \{+, *, a\}, P, S)$, ahol a P szabályai:

$$S \rightarrow A$$

$$A \rightarrow A + B \mid B$$

$$B \rightarrow C * B \mid C$$

$$C \rightarrow a$$

L mátrix:

	S	A	B	C	$+$	$*$	a
S	0	1	0	0	0	0	0
A	0	1	1	0	0	0	0
B	0	0	0	1	0	0	0
C	0	0	0	0	0	0	1

R mátrix:

	S	A	B	C	$+$	$*$	a
S	0	1	0	0	0	0	0
A	0	0	1	0	0	0	0
B	0	0	1	1	0	0	0
C	0	0	0	0	0	0	1

A \triangleleft , \doteq és \triangleright relációk meghatározásában szükségünk lesz ezen relációk tranzitív lezártjára. A tranzitív lezárt könnyen megkapható a

már ismertetett Warshall-algortimussal. A szokásos jelölés a tranzitív lezárt mátrixára: L^+ és R^+ .

L^+ mátrix:

	S	A	B	C	$+$	$*$	a
S	0	1	1	1	0	0	1
A	0	1	1	1	0	0	1
B	0	0	0	1	0	0	1
C	0	0	0	0	0	0	1

R^+ mátrix:

	S	A	B	C	$+$	$*$	a
S	0	1	1	1	0	0	1
A	0	0	1	1	0	0	1
B	0	0	1	1	0	0	1
C	0	0	0	0	0	0	1

Ezen mátrixok segítségével, a relációk értelmezése alapján elkészíthetjük az elsőbbségi (precedencia) táblázatot.

ELőször beírjuk az összes \doteq relációt, megvizsgálva minden szabály jobb oldalát. Az egymást követő betűk között lesz \doteq reláció.

Majd a $\$ \ll X$ meghatározására megnézzük az L^+ mátrixban az S -nek megfelelő sort. Az A, B, C és a oszlopában van egy-egy 1-es. Ezért az elsőbbségi mátrix $\$$ -nak megfelelő sorába beírunk egy-egy \ll jelet az A, B, C és a oszlopokba.

Hasonlóképpen járunk el az $X \gg \$$ meghatározásban, csak itt az R^+ mátrix segítségével dolgozunk.

Az $X \ll Y$ típusú relációknál megkeressük a szabályok jobb oldalán a szomszédos betűket, például $+B$. Megnézzük az L^+ mátrix B -nek megfelelő sorát, itt a C és a oszlopokban van egy-egy 1-es, ezért az elsőbbségi táblázat $+$ jelnek megfelelő sorába beírunk egy-egy \ll jelet a C és a oszlopokba.

Az eredmény a következő táblázatban látható:

	S	A	B	C	$+$	$*$	a	$\$$
S								
A					\doteq			\triangleright
B					\triangleright			\triangleright
C					\triangleright	\doteq		\triangleright
$+$			\doteq	\triangleleft			\triangleleft	
$*$			\doteq	\triangleleft			\triangleleft	
a					\triangleright	\triangleright		\triangleright
$\$$		\triangleleft	\triangleleft	\triangleleft			\triangleleft	

Szerencsénkre a fenti táblázatban minden betűpár között legfőbb egy reláció van. Az üres mező jelzi mindhárom reláció hiányát. Az ilyen nyelvtanok alkalmasak arra, hogy hatékonyan felismerhessük a szvaikat egy alulról felfelé történő elemzéssel.

Egy környezetfüggetlen nyelvtant **egyszerű elsőbbségi (precedencia) nyelvtannak** hívunk, ha

- ε -mentes és nincs olyan változója, amelyre $A \xrightarrow{+} A$,
- a $NUTU\{\$\}$ halmaz bármely két eleme között a \triangleleft , \doteq és \triangleright relációk közül legfeljebb egy áll fenn,
- a helyettesítési szabályok jobb oldalai különböznek egymástól.

Egy egyszerű elsőbbségi nyelvtan segítségével könnyen lehet elemezni, hogy egy szó eleme a nyelvtan által generált nyelvnek vagy sem.

Megvizsgáljuk a szó szomszédos betűi közötti relációkat, minden esetben (ha a szó felismerhető a nyelvtan által) van olyan része, amely \triangleleft jellel kezdődik, utána esetleg \doteq jelek következnek, majd a végén \triangleright van (ez a **nyél**). A nyél mindig egy szabály jobb oldala,

és ez redukálható a szabály bal oldalára. Ezt folytatva, ha a szó eleme a nyelvnek, el lehet jutni a kezdőszimbólumhoz.

Nézzük ezt meg a következő példán! Az $a + a * a$ szót vizsgáljuk.

$\$ \langle a \rangle + a * a \$$	redukálás $C \rightarrow a$
$\$ \langle C \rangle + a * a \$$	redukálás $B \rightarrow C$
$\$ \langle B \rangle + a * a \$$	redukálás $A \rightarrow B$
$\$ \langle A \rangle + \langle a \rangle * a \$$	redukálás $C \rightarrow a$
$\$ \langle A \rangle + \langle C \rangle * \langle a \rangle \$$	redukálás $C \rightarrow a$
$\$ \langle A \rangle + \langle C \rangle * \langle C \rangle \$$	redukálás $B \rightarrow C$
$\$ \langle A \rangle + \langle C \rangle * \langle B \rangle \$$	redukálás $B \rightarrow C * B$
$\$ \langle A \rangle + \langle B \rangle \$$	redukálás $A \rightarrow A + B$
$\$ \langle A \rangle \$$	redukálás $S \rightarrow A$

\$ S \$

Ez megfelel a következő (legjobboldalibb) levezetésnek:

$$\begin{aligned} S &\Rightarrow A \Rightarrow A + B \Rightarrow A + C * B \Rightarrow A + C * C \Rightarrow A + C * a \Rightarrow A + a * a \\ &\Rightarrow B + a * a \Rightarrow C + a * a \Rightarrow a + a * a \end{aligned}$$