

# Fordítóprogramok

*Készítette: Nagy Krisztián*

## Reguláris kifejezések (FLEX)

### Alapelemek kiválasztása

|          |  |
|----------|--|
| $x$      | az 'x' karakter  |
| .        | tetszőleges karakter (kivéve újsor)  |
| [xyz]    | karakterhalmaz; vagy egy 'x', vagy egy 'y' vagy egy 'z'  |
| [abj-oZ] | karakterhalmaz intervallummal  |
| [^A-Z]   | negált karakterhalmaz; bármilyen karakter kivéve nagy betűk  |
| [^A-Z\n] | bármely karakter, kivéve nagybetűk és újsor  |
| \x       | a speciális karakterek, pl. \n, \t; továbbá reguláris kifejezésekben műveletjelek karakterként való használata, pl. \*, \+ |
| \0       | a 0 kódú ASCII karakter  |
| \123     | az oktális 123 kódú karakter   |
| \x2a     | a hexadecimális 2a kódú karakter   |

### Műveletek kiválasztása

|            |   |
|------------|---|
| rs         | r után egy s (konkatenáció)                               |
| $r s$      | egy r vagy egy s (unió)                                   |
| $r^*$      | nulla vagy több r   |
| $(r)$      | r; a zárójelek a műveleti sorrend jelölésére használhatók |
| $r^+$      | egy vagy több r   |
| $r^?$      | nulla vagy egy r (opcionális r)                           |
| $r\{2,5\}$ | kettőtől ötig valahány r                                  |
| $r\{2,\}$  | kettő vagy több r   |
| $r\{4\}$   | pontosan négy r   |
| $^r$       | egy r a sor elején  |
| $r\$$      | egy r a sor végén   |

(A műveletekben r és s tetszőlegesen összetett reguláris kifejezések lehetnek.)

{name} deklarációs részben megadott név kifejtése

## Feladatok (Reguláris kifejezéssel)

1. egész szám (legalább egy számjegy 0-9-ig)
2. olyan egész szám, amely több számjegy esetén nem kezdődhet nullával.
3. előjeles egész szám (opcionálisan + vagy - az elején)
4. törtszám (tizedespont előtt legalább egy számjegy)

## Megoldás

1.  $[0 - 9]^+$
2.  $0|[1 - 9][0 - 9]^*$
3.  $[+ -]?(0|[1 - 9][0 - 9]^*)$
4.  $[0 - 9]^+\backslash.[0 - 9]^+$

## Lexikális elemzés

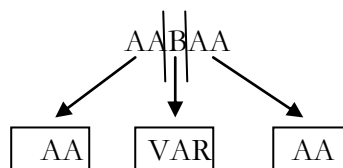
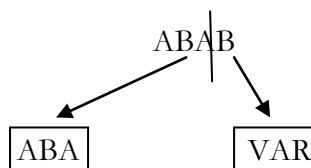
**1. feladat:** Adottak az alábbi lexikai elemek egy-egy reguláris kifejezéssel. Elemezzük ez alapján az ABAB és az AABAA inputokat.

$[a - zA - Z] \rightarrow \boxed{VAR}$

"AB"  $\rightarrow \boxed{AB}$

"ABA"  $\rightarrow \boxed{ABA}$

"AA"  $\rightarrow \boxed{AA}$



**2. feladat:** Adottak az alábbi lexikai elemek egy-egy reguláris kifejezéssel. Elemezzük ez alapján az ALMA123 BEGINEND BEGIN 12 inputot.

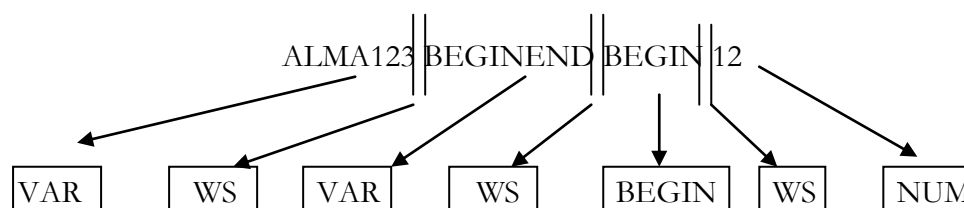
*BEGIN*  $\rightarrow \boxed{BEGIN}$

*END*  $\rightarrow \boxed{END}$

$[a - zA - Z][a - zA - Z0 - 9]^* \rightarrow \boxed{VAR}$

$[0 - 9]^+ \rightarrow \boxed{NUM}$

$[\backslash t \backslash n] \rightarrow \boxed{WS}$



A lexikális elemző betűnként eleméz. Amennyiben egyértelmű a felismerés, akkor a leghosszabb illesztkezés alapján dönti el, hogy melyik lexikai elemre fordítsa az adott inputot a fordító. Abban az esetben, ha nem egyértelmű a felismerés, akkor a lexikai elemek felsorolásában a sorrend alapján dönti el, hogy melyiket válassza. Abban az esetben, ha a megadott inputot nem lehet tokenekre bontani, a fordítás kimenetele lexikális hiba lesz.

## Szintaktikus elemzés

A szintaktikus szabályok leírása kettes típusú (környezetfüggetlen) nyelvtanokkal történik.

### Felülről lefele haladó szintaktikus elemzők

#### LL(1) elemzés

| Egyszerű $LL(1)$  | $\varepsilon$ -mentes $LL(1)$   | Általános $LL(1)$   |
|---|---|---|
| $\varepsilon$ -mentes   | $\varepsilon$ -mentes   | $A \rightarrow a_1$ és $A \rightarrow a_2$ külön szabályok esetén<br>$FIRST_1(\beta_1 FOLLOW_1(A)) \cap FIRST_1(\beta_2 FOLLOW_1(A)) = \emptyset$ |
| szabályok jobboldala terminális szimbólummal kezdődik ( $A \rightarrow a\beta$ alakú)   | $A \rightarrow a_1$ és $A \rightarrow a_2$ külön szabályok esetén<br>$FIRST_1(\beta_1) \cap FIRST_1(\beta_2) = \emptyset$ |   |
| az azonos nemterminálisokhoz tartozó szabályok jobb oldalai különböző terminálissal kezdődnek ( $A \rightarrow a_1\beta$ és $A \rightarrow a_2\beta$ , akkor $a_1 \neq a_2$ ) |   |   |

**1. feladat:** Adottak az alábbi grammatikák. Döntsük el róluk, hogy egyszerű LL(1)-ek –e !  
Amennyiben nem azok indokoljuk meg, hogy miért nem!

- a)  $S \rightarrow aSA|bA$   
 $A \rightarrow \varepsilon|bA$
- b)  $S \rightarrow aSA|A$   
 $A \rightarrow bA|b$
- c)  $S \rightarrow aSA|bA$   
 $A \rightarrow a$

#### **Megoldás:**

- a) Nem egyszerű LL(1), mivel tartalmaz  $A \rightarrow \varepsilon$  szabályt. (Nem  $\varepsilon$ -mentes)
- b) Nem egyszerű LL(1), mivel tartalmaz olyan szabályt, amelynek a jobb oldala nem terminálissal kezdődik. Lásd:  $S \rightarrow A$
- c) Egyszerű LL(1), mivel teljesíti az összes feltételt.

**2. feladat:** Elemezzük a fentebbi feladatban megadott egyszerű **LL(1)** grammatika alapján, hogy az **aabaaa** input szintaktikusan helyes-e elemzővermek segítségével. Írjuk fel a fentebbi grammatika elemzőtáblázatát is.

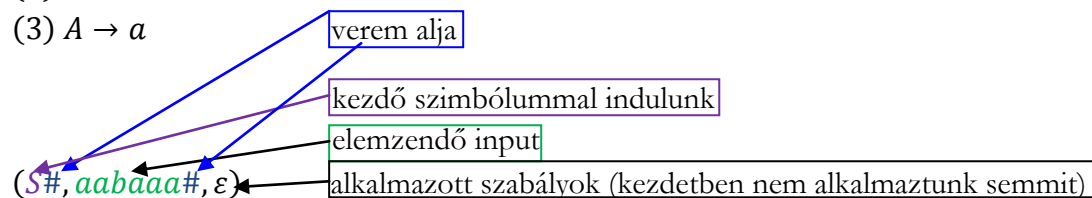
**Megoldás:**

Az egyszerűség kedvéért jelöljük meg a szabályokat számokkal:

(1)  $S \rightarrow aSA$

(2)  $S \rightarrow bA$

(3)  $A \rightarrow a$



$(S\#, aabaaa\#, \varepsilon)$

$(aSA\#, aabaaa\#, 1)$

$(SA\#, abaaa\#, 1)$

$(aSAA\#, abaaa\#, 11)$

$(SAA\#, baaa\#, 11)$

$(bAAA\#, baaa\#, 112)$

$(AAA\#, aaa\#, 112)$

$(aAA\#, aaa\#, 1123)$

$(AA\#, aa\#, 1123)$

$(aA\#, aa\#, 11233)$

$(A\#, a\#, 11233)$

$(a\#, \#, 112333)$

$(\#, \#, 112333)$

Mivel a 2 verem egyszerre lett üres, így sikeres volt az elemzés. Az adott input megfelelő volt.

Elemző táblázat:

Az elemzőtáblázat első oszlopába kezdetben a grammatikában található nem terminálisok kerülnek felsorolásra, majd a terminálisok, végezetül pedig a #.

Az első sorban pedig a grammatikában található terminálisok szerepelnek és végezetül a #.

A táblázat további celláiba az aktuális terminálishoz tartozó szabály kerül. Az üresen maradt cellák pedig hibát jelentenek.

|      | $a$                 | $b$                | $\#$ |
|------|---------------------|--------------------|------|
| $S$  | $S \rightarrow aSA$ | $S \rightarrow bA$ |      |
| $A$  | $A \rightarrow a$   |                    |      |
| $a$  | POP                 |                    |      |
| $b$  |                     | POP                |      |
| $\#$ |                     |                    | OK   |

**3. feladat:** Adottak az alábbi grammatikák. Döntsük el róluk, hogy  $\varepsilon$ -mentes  $LL(1)$ -ek –e !  
Amennyiben nem azok indokoljuk meg, hogy miért nem!

- a)  $S \rightarrow aSA|A$   
 $A \rightarrow bA|a$
- b)  $S \rightarrow SA|Bc$   
 $A \rightarrow aA|c$   
 $B \rightarrow b|c$
- c)  $S \rightarrow dSdS|A|eB$   
 $A \rightarrow a|BA$   
 $B \rightarrow b|c$

**Megoldás:**

- a) Nézzük meg az azonos bal oldalhoz tartozó jobb oldalak FIRST halmazát:

$$FIRST(aSA) = \{a\}$$

$$FIRST(A) = FIRST(bA) \cup FIRST(a) = \{b, a\}$$

$$FIRST(aSA) \cap FIRST(A) = \{a\} \cap \{b, a\} = \{a\} \neq \emptyset$$

Ezért a megadott grammatika NEM  $\varepsilon$ -mentes  $LL(1)$ .

- b) Az  $S \rightarrow SA$  szabály miatt bal rekurzív nyelvtanról beszélhetünk. Ezeket a nyelvtanokat pedig kizárjuk a vizsgálat alól. Ezért a megadott grammatika NEM  $\varepsilon$ -mentes  $LL(1)$ .

- c) Nézzük meg az azonos bal oldalhoz tartozó jobb oldalak FIRST halmazát:

$$FIRST(dSdS) = \{d\}$$

$$FIRST(A) = FIRST(a) \cup FIRST(BA) = \{a, b, c\}$$

$$FIRST(eB) = \{e\}$$

$$FIRST(a) = \{a\}$$

$$FIRST(BA) = FIRST(bA) \cup FIRST(cA) = \{b, c\}$$

$$FIRST(b) = \{b\}$$

$$FIRST(c) = \{c\}$$

Nézzük a metszeteket:

$$FIRST(dSdS) \cap FIRST(A) \cap FIRST(eB) = \{d\} \cap \{a, b, c\} \cap \{e\} = \emptyset$$

$$FIRST(a) \cap FIRST(BA) = \{a\} \cap \{b, c\} = \emptyset$$

$$FIRST(b) \cap FIRST(c) = \{b\} \cap \{c\} = \emptyset$$

Mivel a nyelv  $\varepsilon$ -mentes és az azonos bal oldalhoz tartozó szabályok FIRST halmazainak a metszetei diszjunktak, ezért a megadott grammatika  $\varepsilon$ -mentes  $LL(1)$ .

**4. feladat:** Elemezzük a fentebbi feladatban megadott  $\varepsilon$ -mentes **LL(1)** grammatika alapján, hogy a **dadc** input szintaktikusan helyes-e elemzővermek segítségével. Írjuk fel a fentebbi grammatika elemzőtáblázatát is.

**Megoldás:**

Elemző táblázat:

A táblázatot a FIRST halmazok alapján töltjük ki!

|          | <i>a</i>          | <i>b</i>           | <i>c</i>           | <i>d</i>             | <i>e</i>           | #  |
|----------|-------------------|--------------------|--------------------|----------------------|--------------------|----|
| <i>S</i> | $S \rightarrow A$ | $S \rightarrow A$  | $S \rightarrow A$  | $S \rightarrow dSdS$ | $S \rightarrow eB$ |    |
| <i>A</i> | $A \rightarrow a$ | $A \rightarrow BA$ | $A \rightarrow BA$ |                      |                    |    |
| <i>B</i> |                   | $B \rightarrow b$  | $B \rightarrow c$  |                      |                    |    |
| <i>a</i> | POP               |                    |                    |                      |                    |    |
| <i>b</i> |                   | POP                |                    |                      |                    |    |
| <i>c</i> |                   |                    | POP                |                      |                    |    |
| <i>d</i> |                   |                    |                    | POP                  |                    |    |
| <i>e</i> |                   |                    |                    |                      | POP                |    |
| #        |                   |                    |                    |                      |                    | OK |

Elemzés vermekkel:

$(S\#, dadc\#, \varepsilon)$   
 $(dSdS\#, dadc\#, 1)$   
 $(SdS\#, adc\#, 1)$   
 $(AdS\#, adc\#, 12)$   
 $(aS\#, adc\#, 124)$   
 $(dS\#, dc\#, 124)$   
 $(S\#, c\#, 124)$   
 $(A\#, c\#, 1242)$   
 $(BA\#, c\#, 12425)$   
 $(cA\#, c\#, 124257)$

Szintaktikai hiba, mivel az elemzendő inputot tartalmazó verem kiürült úgy, hogy a másik verem még nem.

**5. feladat:** Adott az alábbi grammatika. Döntsük el róla, hogy általános LL(1)-e!

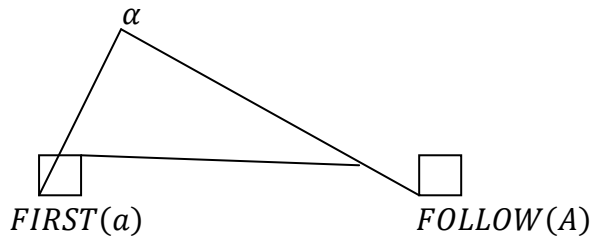
$$S \rightarrow [A]$$

$$A \rightarrow \varepsilon | nB$$

$$B \rightarrow \varepsilon | , nB$$

$$FIRST(\varepsilon) = ?$$

$A \rightarrow \alpha$  típusú szabályok esetén:



$FOLLOW(A)$  - Mi az első olyan terminális elem, ami nem  $\alpha$ -ból lett levezetve, de az első utána.

$$FIRST([A] FOLLOW(S)) = \{\}$$

$$FIRST(\varepsilon FOLLOW(A)) = FOLLOW(A) = \{\}$$

$$FOLLOW(A) = FIRST(]FOLLOW(S)) = \{\}$$

$$FIRST(nB FOLLOW(A)) = \{n\}$$

$$FIRST(\varepsilon FOLLOW(B)) = FOLLOW(B) = \{\}$$

$$\begin{aligned} FOLLOW(B) &= FIRST(\varepsilon FOLLOW(A)) \cup FIRST(\varepsilon FOLLOW(B)) = \\ &= FOLLOW(A) \cup FOLLOW(B) = \{\} \cup FOLLOW(B) \end{aligned}$$

Megjegyzés: Folyamatosan ezt állítja elő  $FOLLOW(B)$ , így az  $\cup$ -ba nem kerül új elem!

$$FIRST(, nB FOLLOW(B)) = \{, \}$$

Mivel  $FIRST(\varepsilon FOLLOW(A)) \cap FIRST(nB FOLLOW(A)) = \emptyset$  és

$FIRST(\varepsilon FOLLOW(B)) \cap FIRST(, nB FOLLOW(B)) = \emptyset$  ezért a fentebb megadott grammatika általános LL(1)

**6. feladat:** Írjuk fel az előző feladatban látott általános LL(1) grammatika elemző táblázatát.

Az elemzőtáblázatot a FIRST halmazok alapján írjuk fel.

|          | [                   | ]                           | ,                   | <i>n</i>           | #  |
|----------|---------------------|-----------------------------|---------------------|--------------------|----|
| <i>S</i> | $S \rightarrow [A]$ |                             |                     |                    |    |
| <i>A</i> |                     | $A \rightarrow \varepsilon$ |                     | $A \rightarrow nB$ |    |
| <i>B</i> |                     | $B \rightarrow \varepsilon$ | $B \rightarrow ,nB$ |                    |    |
| [        | POP                 |                             |                     |                    |    |
| ]        |                     | POP                         |                     |                    |    |
| ,        |                     |                             | POP                 |                    |    |
| <i>n</i> |                     |                             |                     | POP                |    |
| #        |                     |                             |                     |                    | OK |

### Alulról felfele haladó szintaktikus elemzők

#### LR(0) elemzés

**Feladat:** Adott az alábbi grammatika. Lássuk be  $LR(0)$  elemzés segítségével, hogy az **axbx** mondata- e a nyelvnek. Írjuk fel a kannonikus halmazokat, készítsünk automatát és ellenőrizzük veremmel a fentebbi kérdést!

$$S \rightarrow x|AS$$

$$A \rightarrow aSb$$

**Megoldás:**

Egészítsük ki a nyelvtant:

$$S' \rightarrow S$$

$$S \rightarrow x|AS$$

$$A \rightarrow aSb$$

#### Kannonikus halmazok

LR(0) elem:  $[A \rightarrow .\alpha]$ , ahol a  $.$  jelzi, hogy hol tartunk adott input beolvasása után az addig elemzett inputtal.

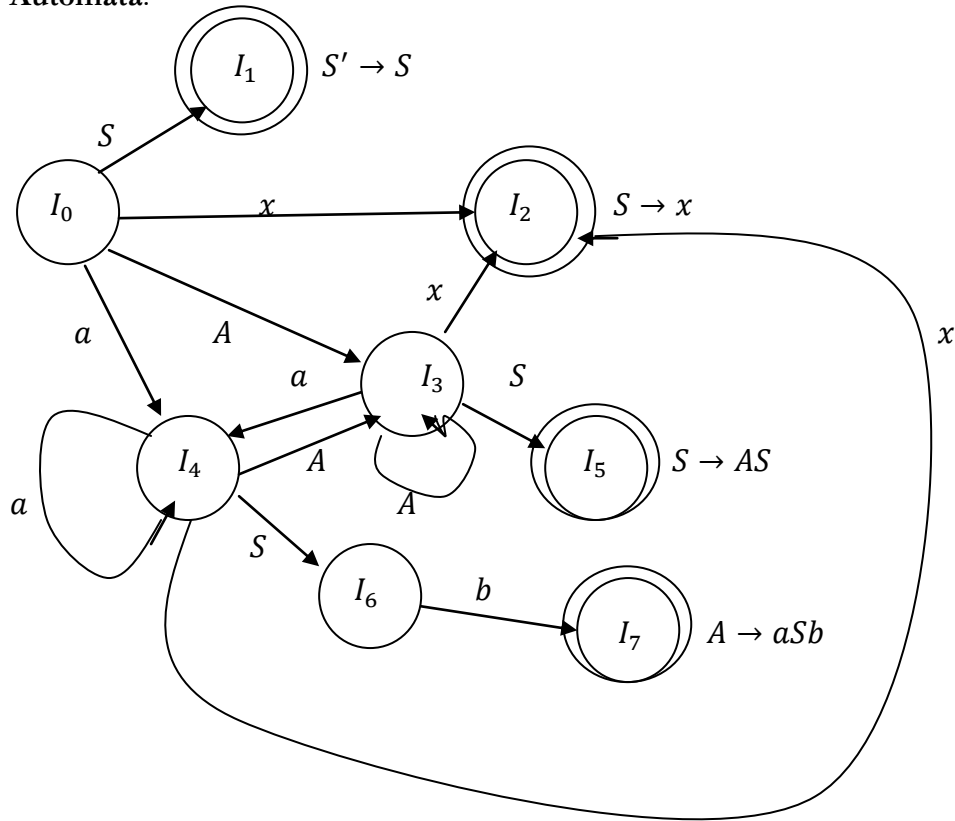
$$I_0 = CLOSURE([S' \rightarrow .S]) = \{[S' \rightarrow .S], [S \rightarrow .x], [S \rightarrow .AS], [A \rightarrow .aSb]\}$$

Megjegyzés: Minden CLOSURE halmazt addig bővítünk, ameddig képes bővülni. Az aktuális szabály is a halmaz része. Akkor fejeződött be a teljes feldolgozás, ha az összes pont a szabályok jobb oldalán a szabályok végére kerül.



$I_0 = \text{CLOSURE}([S' \rightarrow \cdot S]) = \{[S' \rightarrow \cdot S], [S \rightarrow \cdot x], [S \rightarrow \cdot AS], [A \rightarrow \cdot aSb]\}$   
 $I_1 = \text{READ}(I_0, S) = \text{CLOSURE}([S' \rightarrow S \cdot]) = \{[S' \rightarrow S \cdot]\}$   
 $I_2 = \text{READ}(I_0, x) = \text{CLOSURE}([S \rightarrow x \cdot]) = \{[S \rightarrow x \cdot]\}$   
 $I_3 = \text{READ}(I_0, A) = \text{CLOSURE}([S \rightarrow A \cdot S]) = \{[S \rightarrow A \cdot S], [S \rightarrow \cdot x], [S \rightarrow \cdot AS], [A \rightarrow \cdot aSb]\}$   
 $I_4 = \text{READ}(I_0, a) = \text{CLOSURE}([A \rightarrow a \cdot Sb]) =$   
 $= \{[A \rightarrow a \cdot Sb], [S \rightarrow \cdot x], [S \rightarrow \cdot AS], [A \rightarrow \cdot aSb]\}$   
 $I_5 = \text{READ}(I_3, S) = \text{CLOSURE}([S \rightarrow AS \cdot]) = \{[S \rightarrow AS \cdot]\}$   
 $\text{READ}(I_3, x) = \text{CLOSURE}([S \rightarrow x \cdot]) \equiv I_2$   
 $\text{READ}(I_3, A) = \text{CLOSURE}([S \rightarrow A \cdot S]) \equiv I_3$   
 $\text{READ}(I_3, a) = \text{CLOSURE}([A \rightarrow a \cdot Sb]) \equiv I_4$   
 $I_6 = \text{READ}(I_4, S) = \text{CLOSURE}([A \rightarrow aS \cdot b]) = \{[A \rightarrow aS \cdot b]\}$   
 $\text{READ}(I_4, x) = \text{CLOSURE}([S \rightarrow x \cdot]) \equiv I_2$   
 $\text{READ}(I_4, A) = \text{CLOSURE}([S \rightarrow A \cdot S]) \equiv I_3$   
 $\text{READ}(I_4, a) = \text{CLOSURE}([A \rightarrow a \cdot Sb]) \equiv I_4$   
 $I_7 = \text{READ}(I_6, b) = \text{CLOSURE}([A \rightarrow aSb \cdot]) = \{[A \rightarrow aSb \cdot]\}$

**Automata:**



### Elemzés veremmel:

$(axbx\#, \#0)$  - 2. paraméter : állapot

$(xbx\#, \#0a4)$

$(bx\#, \#0a4x2)$

$(bx\#, \#0a4S6)$

$(x\#, \#0a4S6b7)$

$(x\#, \#0A3)$

$(\#, \#0A3x2)$

$(\#, \#0A3S5)$

$(\#, \#0S1)$

$(\#, \#0S')$

Megjelent  $S'$  helyesen, így az adott  $axbx$  kifejezés mondat a nyelvben.

### Elemző táblázat:

|   | action    | $S'$                | $S$ | $A$ | $x$ | $a$ | $b$ |
|---|-----------|---------------------|-----|-----|-----|-----|-----|
| 0 | shift     |                     | 1   | 3   | 2   | 4   |     |
| 1 | reduction | $S' \rightarrow S$  |     |     |     |     |     |
| 2 | reduction | $S \rightarrow x$   |     |     |     |     |     |
| 3 | shift     |                     | 5   | 3   | 2   | 4   |     |
| 4 | shift     |                     | 6   | 3   | 2   | 4   |     |
| 5 | reduction | $S \rightarrow AS$  |     |     |     |     |     |
| 6 | shift     |                     |     |     |     |     | 7   |
| 7 | reduction | $A \rightarrow aSb$ |     |     |     |     |     |

Amire figyelni kell:

$LR(0)$  egy nyelvtan, ha konfliktus mentes a kannonikus alakja.

Lehetséges konfliktusok:

Shift/Shift  $\rightarrow$  egy állapotból 2 állapotba lehet eljutni ugyan azzal a szimbólummal

Reduction/Reduction  $\rightarrow$  „két szabály a táblázat valamelyik redukáló sorában”

Shift/Reduction  $\rightarrow$  redukáló állapot, de vezet ki belőle él.