

## Nyelvtani transzformációk

### Formális nyelvek, 6. gyakorlat

**Célja:** A nyelvtani transzformációk bemutatása

**Fogalmak:** Megszorított típusok, normálformák, 0. típusú epsilon-mentesítés, 2. típusú epsilon-mentesítés, láncmentesítés, Chomsky-féle normálformává alakítás, reguláris műveletekre való zártsgot bizonyító konstrukciók.

**Feladatok jellege:** Konkrét nyelvtanokból kiindulva a konstrukciók tényleges, mechanikus elvégzése.

2008/09 I. félév

## Házi feladatok megoldása

### 1. feladat

Melyik nyelvet generálja a következő nyelvtan?  $T = \{(\cdot)\}$

a.  $S \rightarrow (S) \mid SS \mid \varepsilon$

b.  $S \rightarrow XS \mid \varepsilon$  és  $X \rightarrow (S)$

c.  $S \rightarrow (SS \mid \varepsilon)$

### Megoldás:

a. és b.: HE, c.: HE). Például a:

HE, a helyes zárójelezések nyelve, azaz  $u \in \{(\cdot)\}^* \in \text{HE} \Leftrightarrow \ell_1(u) = \ell_2(u)$  és minden  $v \in \text{Pre}(u)$  esetén  $\ell_1(v) \geq \ell_2(v)$ .

Jelölje  $L$  a generált nyelvet.

" $L \subseteq \text{HE}$ ":

Minden újonnan behozott jobbzárójel elé valahova kerül ugyanakkor egy új balzárójel is. Így az aktuális mondatforma minden prefixében legalább annyi balzárójel van mint jobb. Formálisan:

## Házi feladatok megoldása

### 1. feladat

Melyik nyelvet generálja a következő nyelvtan?  $T = \{(\cdot)\}$

a.  $S \rightarrow (S) \mid SS \mid \varepsilon$

b.  $S \rightarrow XS \mid \varepsilon$  és  $X \rightarrow (S)$

c.  $S \rightarrow (SS \mid \varepsilon)$

Ha  $\alpha \in \{S, (\cdot)\}^*$ , legyen  $\phi(\alpha)$  a következő tulajdonság:

$\forall u \in \text{Pre}(\alpha) : \ell_1(u) \geq \ell_2(u)$  és  $\ell_1(\alpha) = \ell_2(\alpha)$ .

Állítás: Ha  $S \xrightarrow{*} \alpha$ , akkor  $\phi(\alpha)$ .

A levezetés hosszára vonatkozó teljes indukcióval bizonyítunk. Ha  $\alpha = S$ , akkor igaz az állítás.

Tegyük fel, hogy  $S \xrightarrow{n} \alpha = \alpha_1 S \alpha_2 \rightarrow \beta_i$  ( $i = 1, 2, 3$ ), ahol  $\beta_1 = \alpha_1 (S) \alpha_2$ ,  $\beta_2 = \alpha_1 SS \alpha_2$ ,  $\beta_3 = \alpha_1 \alpha_2$ .

Az indukciós feltevés miatt  $\phi(\alpha)$ , be kell látni, hogy  $\phi(\beta_i)$  ( $i = 1, 2, 3$ ).

Legyen  $u \in \text{Pre}(\beta_i)$ , azaz  $uv = \beta_i$  ( $i = 1, 2, 3$ ).

## Házi feladatok megoldása

### 1. feladat

Melyik nyelvet generálja a következő nyelvtan?  $T = \{(\cdot)\}$

a.  $S \rightarrow (S) \mid SS \mid \varepsilon$

b.  $S \rightarrow XS \mid \varepsilon$  és  $X \rightarrow (S)$

c.  $S \rightarrow (SS \mid \varepsilon)$

Ha  $u \in \text{Pre}(\alpha_1)$ , akkor az indukció alapján  $\ell_1(u) \geq \ell_2(u)$ .

Ha  $v \in \text{Suf}(\alpha_2)$ , akkor  $\exists u', u'v = \alpha$ . Az indukció alapján  $\ell_1(u') \geq \ell_2(u')$  és akármelyik szabályt is alkalmaztuk, ugyanannyival ( $i = 2, 3$  esetén 0-val,  $i = 1$  esetén 1-gyel) nőtt a bal- és jobbzárójelek száma  $u$ -ban  $u'$ -höz képest. Tehát  $\ell_1(u) \geq \ell_2(u)$ .

Ha  $u = \alpha_1 \gamma_1$  és  $v = \gamma_2 \alpha_2$ , (azaz  $\gamma_1$  prefixe valamelyik levezetési szabály jobboldalának) akkor indukció alapján  $\ell_1(\alpha_1) \geq \ell_2(\alpha_1)$  és könnyen ellenőrizhető, hogy  $\ell_1(\gamma_1) \geq \ell_2(\gamma_1)$ , tehát  $\ell_1(u) = \ell_1(\alpha_1) + \ell_1(\gamma_1) \geq \ell_2(\alpha_1) + \ell_2(\gamma_1) = \ell_2(u)$ .

$\ell_1(\beta_i) = \ell_2(\beta_i)$ , hiszen minden szabály jobboldala ugyanannyi bal- és jobbzárójelet tartalmaz (0-t vagy 1-et), tehát  $\phi(\beta_i)$ , ( $i = 1, 2, 3$ ).

## Házi feladatok megoldása

### 1. feladat

Melyik nyelvet generálja a következő nyelvtan?  $T = \{ (, ) \}$

a.  $S \rightarrow (S) | SS | \varepsilon$       b.  $S \rightarrow XS | \varepsilon$  és  $X \rightarrow (S)$       c.  $S \rightarrow (SS)$

" $L \supseteq HE$ ":

A zárójelek számára vonatkozó indukcióval belátjuk, hogy minden helyes zárójelezés levezethető.

Az üres zárójelezés az  $S \rightarrow \varepsilon$  szabállyal levezethető.

Tekintsünk egy  $w$  helyes zárójelezést. Ekkor vagy  $w = (w_1)$ , vagy  $w = w_1 w_2$ , ahol  $w_1, w_2$  helyes zárójelezések. (Attól függően, hogy van-e  $w$ -nek valódi prefixe, mely ugyanannyi bal- és jobbzárójelet tartalmaz.)

Indukció alapján  $w_1$  és  $w_2$  levezethető.

Az első esetben az  $S \rightarrow (S) \xrightarrow{*} (w_1)$ , a másodikban az

$S \rightarrow SS \xrightarrow{*} w_1 S \xrightarrow{*} w_1 w_2$  levezetés  $w$ -nek egy jó levezetését adja.

## Házi feladatok megoldása

### 2. feladat

Adjunk az  $L = \{v; v = uu\}$  ( $T = \{a, b\}$ ) nyelvet generáló nyelvtant a " $vv^{-1}$ " alakú szavak nyelvénél látott módszerre való visszavezetéssel!

$$\begin{aligned} S' &\rightarrow \varepsilon \mid S \\ S &\rightarrow tSX_t \mid tY_t \quad \forall t \in T \\ Y_t X_{t'} &\rightarrow Y_{t'} t \quad \forall t, t' \in T \\ tX_{t'} &\rightarrow X_{t'} t \quad \forall t, t' \in T \\ Y_t &\rightarrow t \end{aligned}$$

## Házi feladatok megoldása

### 3. feladat

Készítsünk olyan 3. típusú  $G$  nyelvtant, melyre  $L(G) = (b(ab \cup a)^* \cup ab)^* \cup aa^*$

$$\begin{aligned} S &\rightarrow \varepsilon \mid A \mid aa \\ A &\rightarrow B \mid ab \mid abA \\ B &\rightarrow bC \\ C &\rightarrow \varepsilon \mid B \mid D \\ D &\rightarrow ab \mid a \mid abD \mid aD \mid abA \mid aA \end{aligned}$$

## Normálformák

Típus	Normálforma szabályai	Lépések
0.	$S \rightarrow \varepsilon$ , ez esetben $S$ nem szerepel szabály jobboldalán $p \rightarrow q, q \in (T \cup N)^+$	0. típusú $\varepsilon$ -mentesítés
1. (Kuroda)	$S \rightarrow \varepsilon$ , ez esetben $S$ nem szerepel szabály jobboldalán $A \rightarrow a$ $A \rightarrow BC$ $AB \rightarrow AC$ $BA \rightarrow CA$	álterminálisok hosszredukció $AB \rightarrow CD, A \neq C, B \neq D$ alakúak eliminálása
2. (Greibach)	$S \rightarrow \varepsilon$ , ez esetben $S$ nem szerepel szabály jobboldalán $A \rightarrow aQ, Q \in N^*$	redukció rendezetté tétel kvázi Greibach forma Greibach NF-ra hozás

## Normálformák (folyt.)

2. (Chomsky)	$S \rightarrow \varepsilon$ , ez esetben $S$ nem szerepel szabály jobboldalán $A \rightarrow a$ $A \rightarrow BC$	redukció $\varepsilon$ -mentesítés láncmentesítés áletterminálisok hosszredukció
3.	$A \rightarrow \varepsilon$ $A \rightarrow aB$	láncmentesítés hosszredukció $A \rightarrow a$ alakú szabályok eliminálása

## 0. típusú nyelvtanok

$\varepsilon$ -mentesítés

$$G = (\{a, b\}, \{S, X, Y\}, \mathcal{P}, S)$$

$$S \rightarrow aXSbY \mid ab \mid \varepsilon$$

$$Xb \rightarrow \varepsilon$$

$$Xa \rightarrow aaX$$

Végezzük el a 0. típusú  $\varepsilon$ -mentesítést!

Megoldás:

$$S' \rightarrow \varepsilon \mid S$$

$$S \rightarrow aXSbY \mid ab$$

$$aS \rightarrow a \quad Sa \rightarrow a \quad aXb \rightarrow a \quad Xba \rightarrow a$$

$$bS \rightarrow b \quad Sb \rightarrow b \quad bXb \rightarrow b \quad Xbb \rightarrow b$$

$$XS \rightarrow X \quad SX \rightarrow X \quad XXb \rightarrow X \quad XbX \rightarrow X$$

$$YS \rightarrow Y \quad SY \rightarrow Y \quad YXb \rightarrow Y \quad XbY \rightarrow Y$$

$$SS \rightarrow S \quad SXb \rightarrow S \quad XbS \rightarrow S$$

$$Xa \rightarrow aaX$$

## Kuroda normálforma (1. típusú nyelvtan)

A normálformára alakítás lépései

1. lépés: **Átterminális** nyelvtani jelek bevezetése:

Minden terminálist egy áletterminális nyelvtani jellel helyettesítünk, és az áletterminálist terminálisra cserélő szabályokat adunk a rendszerhez.

Példa:  $ABC \rightarrow DEFgH$  szabályból  $ABQ_c \rightarrow DEFQ_gH$  lesz és a  $Q_c \rightarrow c$  valamint a  $Q_g \rightarrow g$  szabályt hozzáadjuk a szabályrendszerhez.

2. lépés: **Hosszredukció**: új, egyedi nyelvtani jelekkel minden hosszú szabályt vele ekvivalens rövid szabályokkal helyettesítünk.

Példa:  $ABQ_c \rightarrow DEFQ_gH$  szabály helyett lesz  $(Z_1, Z_2, Z_3$  új, egyedi jel):

$$AB \rightarrow DZ_1 \quad Q_cZ_1 \rightarrow EZ_2 \quad Z_2 \rightarrow FZ_3 \quad Z_3 \rightarrow Q_gH$$

3. lépés: Az " $AB \rightarrow CD, A \neq C, B \neq D$ " alakú szabályok eliminálása:

Minden szabályt 3, már helyes újjal helyettesítünk.

Példa:  $AB \rightarrow DZ_1$  szabály helyett lesz  $(W$  új, egyedi nyelvtani jel)

$$AB \rightarrow AW \quad AW \rightarrow DW \quad DW \rightarrow DZ_1$$

## Chomsky normálforma (2. típus)

0. lépés: Redukció – Zsákutcamentesítés

$$G = (\{a, b, c\}, \{S, A, B, C, D, E\}, \mathcal{P}, S)$$

$$S \rightarrow AB \mid CD \quad C \rightarrow Cccc \mid \varepsilon$$

$$A \rightarrow aAa \mid C \quad D \rightarrow DD \mid DS$$

$$B \rightarrow bBb \mid C \quad E \rightarrow aS \mid bS$$

**Zsákutcamentesítés**: Azon nyelvtani jelek elhagyása, melyekből nem vezethető le terminális szó. Meghatározzuk azon nyelvtani jelek  $H \subseteq N$  halmazát, melyekből levezethető valamilyen terminális szó.

A  $H_i$  halmaz olyan nyelvtani jelekből áll, melyekből közvetlenül levezethető  $(H_{i-1} \cup T)^*$ -beli mondatforma, azaz legfeljebb  $i$  levezetési fázisban levezethető belőle terminális szó. A  $H_1, H_2, \dots$  halmzsorozat a tartalmazásra nézve monoton növekvő, felülről korlátos halmzsorozat, így mindig stabilizálódik. Ha  $H_{i+1} = H_i$ , akkor legyen  $H = H_i$ .

A  $H_i$  halmaz ismeretében meghatározhatjuk  $H_{i+1}$ -t:  
 $H_{i+1} = H_i \cup \{X \in N \mid \exists q : X \rightarrow q \in \mathcal{P}, q \in (H_i \cup T)^*\}$ .

## Chomsky normálforma (2. típus)

0. lépés: Redukció – Zsákutcamentesítés (folyt.)

$$G = \langle \{a, b, c\}, \{S, A, B, C, D, E\}, \mathcal{P}, S \rangle$$
$$S \rightarrow AB \mid CD \quad C \rightarrow Cccc \mid \varepsilon$$
$$A \rightarrow aAa \mid C \quad D \rightarrow DD \mid DS$$
$$B \rightarrow bBb \mid C \quad E \rightarrow aS \mid bS$$

$$H_1 = \{C\},$$
$$H_2 = \{C, A, B\},$$
$$H_3 = \{C, A, B, S\},$$
$$H_4 = \{C, A, B, S, E\},$$
$$H_5 = H_4, \Rightarrow H = H_4 = \{C, A, B, S, E\}.$$

A  $D$  nyelvtani jel zsákutcába vezet. Elhagyunk minden olyan szabályt, ami tartalmazza a  $D$ -t. (Általában, minden olyat ami tartalmaz  $N \setminus H$ -belit.)

## Chomsky normálforma (2. típus)

0. lépés: Redukció – Összefüggővé alakítás

$$G = \langle \{a, b, c\}, \{S, A, B, C, E\}, \mathcal{P}, S \rangle$$
$$S \rightarrow AB \quad C \rightarrow Cccc \mid \varepsilon$$
$$A \rightarrow aAa \mid C \quad E \rightarrow aS \mid bS$$
$$B \rightarrow bBb \mid C$$

**Összefüggővé alakítás:** Azon nyelvtani jelek elhagyása, melyek nem szerepelhetnek az  $S$  kezdőszimbólumból levezetett mondatformákban. Megkonstruáljuk a levezetések során előforduló nyelvtani jelek  $K \subseteq N$  halmazát.

$K_0 = \{S\}$ . A  $K_i$  halmaz álljon azon nyelvtani jelekből, melyek a legfeljebb  $i$  lépéses levezetések során bejöhethetnek. A  $K_0, K_1, \dots$  halmzsorozat a tartalmazásra nézve monoton növekvő, felülről korlátos halmzsorozat, így mindig stabilizálódik.

$K_{i+1} = K_i \cup \{X \in N \mid \exists Y \rightarrow q \in \mathcal{P} : X \subseteq q\}$ , azaz  $K_{i+1}$  azon nyelvtani jelekből áll, melyek szerepelnek valamely  $K_i$ -beli nyelvtani jelre vonatkozó szabály jobboldalán.

## Chomsky normálforma (2. típus)

0. lépés: Redukció – Összefüggővé alakítás (folyt.)

$$G = \langle \{a, b, c\}, \{S, A, B, C, E\}, \mathcal{P}, S \rangle$$
$$S \rightarrow AB \quad C \rightarrow Cccc \mid \varepsilon$$
$$A \rightarrow aAa \mid C \quad E \rightarrow aS \mid bS$$
$$B \rightarrow bBb \mid C$$

$$K_0 = \{S\},$$
$$K_1 = \{S, A, B\},$$
$$K_2 = \{S, A, B, C\},$$
$$K_3 = K_2, \Rightarrow K = K_2 = \{S, A, B, C\}.$$

Az  $E$  nyelvtani jel nem jöhet be  $S$ -ből való levezetések során. Elhagyunk minden olyan szabályt, mely tartalmazza az  $E$  nyelvtani jelet. (Általában  $N \setminus K$ -belit.)

### Redukált nyelvtan

A zsákutcamentes és összefüggő 2-es típusú nyelvtanokat **redukáltnak** nevezzük.

## Chomsky normálforma (2. típus)

1. lépés:  $\varepsilon$ -mentesítés

$$G = \langle \{a, b, c\}, \{S, A, B, C\}, \mathcal{P}, S \rangle$$
$$S \rightarrow AB$$
$$A \rightarrow aAa \mid C$$
$$B \rightarrow bBb \mid C$$
$$C \rightarrow Cccc \mid \varepsilon$$

### $\varepsilon$ -mentesítés:

Ismét konstruálunk egy  $H \subseteq \{S, A, B, C\}$  segédhalmazt, melynek pontosan azok a nyelvtani jelek lesznek az elemei, melyekből (egy vagy több lépésben) levezethető  $\varepsilon$ . Ehhez ismét rekurzívan definiálunk  $H_i$  halmazokat. A  $H_i$  halmaz a  $H_{i-1}$  halmaz bővítése azon nyelvtani jelekkel, amelyekből közvetlenül levezethető  $H_{i-1}^*$ -beli mondatforma. A kiindulási halmaz  $H_1$ , azon nyelvtani jelek halmaza, melyekből közvetlenül levezethető  $\varepsilon$ .

## Chomsky normálforma (2. típus)

1. lépés:  $\varepsilon$ -mentesítés (folyt.)

$$H_1 = \{C\}, \quad H_3 = \{A, B, C\} \cup \{S\}, \\ H_2 = \{C\} \cup \{A, B\}, \quad H_4 = H_3 \Rightarrow H = \{S, A, B, C\}.$$

Képezzük az összes olyan szabályt, melynek jobboldala valamely eredeti szabály jobboldalán szereplő mondatformából tetszőlegesen kiválasztott  $H$ -beli nemterminálisok elhagyásával kapható (beleértve azt az esetet is, ha nem hagyunk el semmit), a baloldal pedig az eredet baloldal. Az  $\varepsilon$ -szabályokat elhagyjuk.

$$S \rightarrow AB \mid A \mid B \\ A \rightarrow aAa \mid aa \mid C \\ B \rightarrow bBb \mid bb \mid C \\ C \rightarrow Cccc \mid ccc$$

Ha  $S \in H$  (mint az adott példában) akkor hozzá kell adni még a következő szabályokat:

$$S' \rightarrow S \mid \varepsilon \quad (\text{ekkor } S' \text{ lesz az új kezdőszimbólum})$$

## Chomsky normálforma (2. típus)

2. lépés: Láncmentesítés

Láncmentesítés:

$$S' \rightarrow S \mid \varepsilon \\ S \rightarrow AB \mid A \mid B \\ A \rightarrow aAa \mid aa \mid C \\ B \rightarrow bBb \mid bb \mid C \\ C \rightarrow Cccc \mid ccc$$

Meghatározzuk minden nyelvtani jelhez (az adott nyelvtani jelet magát is beleértve) azon nyelvtani jelek halmazát, melyek mint 1 hosszúságú mondatforma (közvetlenül, vagy közvetetten) levezethetők belőle.

Ezeket a halmazokat is rekurzívan határozhatjuk meg. Például:

$$H_0(S) = \{S\}, \\ H_1(S) = H_0(S) \cup \{A, B\} = \{S, A, B\}, \\ H_2(S) = H_1(S) \cup \{C\} = \{S, A, B, C\}, \\ H_3(S) = H_2(S) \Rightarrow H(S) = H_2(S).$$

## Chomsky normálforma (2. típus)

2. lépés: Láncmentesítés (folyt.)

$$S' \rightarrow S \mid \varepsilon \quad H(S') = \{S', S, A, B, C\}, \\ S \rightarrow AB \mid A \mid B \quad H(S) = \{S, A, B, C\}, \\ A \rightarrow aAa \mid aa \mid C \quad H(A) = \{A, C\}, \\ B \rightarrow bBb \mid bb \mid C \quad H(B) = \{B, C\}, \\ C \rightarrow Cccc \mid ccc \quad H(C) = \{C\}.$$

Minden  $Y \in H(X)$  nyelvtani jelhez vesszük azon szabályokat, amelyeknek baloldalán  $X$ , jobboldalán pedig egy  $Y$ -ra vonatkozó eredeti szabály jobboldala áll, kivéve ha ez a jobboldali mondatforma egyetlen nyelvtani jeltől áll. (Azaz, a láncszabályokat elhagyjuk.)

$$G = \langle \{a, b, c\}, \{S', S, A, B, C\}, \mathcal{P}, S' \rangle$$

$$S' \rightarrow \varepsilon \mid AB \mid aAa \mid aa \mid bBb \mid bb \mid Cccc \mid ccc \\ S \rightarrow AB \mid aAa \mid aa \mid bBb \mid bb \mid Cccc \mid ccc \\ A \rightarrow aAa \mid aa \mid Cccc \mid ccc \\ B \rightarrow bBb \mid bb \mid Cccc \mid ccc \\ C \rightarrow Cccc \mid ccc$$

## Chomsky normálforma (2. típus)

3. lépés: Álterminálisok bevezetése, 4. lépés: Hosszredukció

Álterminálisok bevezetése:

A terminálisokat egy-egy terminális szimbólumként egyedi, új nyelvtani jelre cseréljük (kivéve ha a jobboldal már egyetlen terminális szimbólumból áll), és az ezeket a megfelelő terminálisokra cserélő szabályokat hozzáadjuk.

$$S' \rightarrow \varepsilon \mid AB \mid Q_a A Q_a \mid Q_a Q_a \mid Q_b B Q_b \mid Q_b Q_b \mid C Q_c Q_c Q_c \mid Q_c Q_c Q_c \\ S \rightarrow AB \mid Q_a A Q_a \mid Q_a Q_a \mid Q_b B Q_b \mid Q_b Q_b \mid C Q_c Q_c Q_c \mid Q_c Q_c Q_c \\ A \rightarrow Q_a A Q_a \mid Q_a Q_a \mid C Q_c Q_c Q_c \mid Q_c Q_c Q_c \\ B \rightarrow Q_b B Q_b \mid Q_b Q_b \mid C Q_c Q_c Q_c \mid Q_c Q_c Q_c \\ C \rightarrow C Q_c Q_c Q_c \mid Q_c Q_c Q_c \\ Q_a \rightarrow a \quad Q_b \rightarrow b \quad Q_c \rightarrow c$$

Hosszredukció:

Mint a Kuroda NF-nél. Például:  $C \rightarrow C Q_c Q_c Q_c$  szabály helyett:

$$C \rightarrow C Z_1 \quad Z_1 \rightarrow Q_c Z_2 \quad Z_2 \rightarrow Q_c Q_c$$

### 3. típusú nyelvtanok normálformája

$$S \rightarrow \varepsilon \mid bA$$
$$A \rightarrow aaA \mid S \mid b$$

1. lépés: Lánccmentesítés:

$$S \rightarrow \varepsilon \mid bA$$
$$A \rightarrow aaA \mid \varepsilon \mid bA \mid b$$

2. lépés: Hosszredukció:

$$S \rightarrow \varepsilon \mid bA$$
$$A \rightarrow aZ_1 \mid \varepsilon \mid bA \mid b$$
$$Z_1 \rightarrow aA$$

3. lépés: A nyelvtani jelből terminális alakú szabályok átalakítása:

$$S \rightarrow \varepsilon \mid bA$$
$$A \rightarrow aZ_1 \mid \varepsilon \mid bA \mid bF_1$$
$$Z_1 \rightarrow aA$$
$$F_1 \rightarrow \varepsilon$$