Nyelvtani transzformációk

Formális nyelvek, 6. gyakorlat

Célja: A nyelvtani transzformációk bemutatása

Fogalmak: Megszorított típusok, normálformák, 0. típusú epszilon-mentesítés, 2. típusú epszilon-mentesítés, láncmentesítés, Chomsky-féle normálformává alakítás, reguláris műveletekre való zártságot 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? $T = \{(,)\}$
a. $S \to (S) \mid SS \mid \varepsilon$ **b.** $S \to XS \mid \varepsilon$ *és* $X \to (S)$ **c.** $S \to (SS \mid)$

Megoldás:

a. és **b.**: **HE**, **c.**: **HE**). Például **a**:

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

Jelölje L a generált nyelvet.

"$L \subseteq$ **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? $T = \{(,)\}$
a. $S \to (S) \mid SS \mid \varepsilon$ **b.** $S \to XS \mid \varepsilon$ *és* $X \to (S)$ **c.** $S \to (SS \mid)$

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

$\forall u \in \mathrm{Pre}(\alpha): \quad \ell_((u) \geq \ell_)(u)$ és $\quad \ell_((\alpha) = \ell_)(\alpha)$.

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 $S \xrightarrow{n} \alpha = \alpha_1 S \alpha_2 \to \beta_i$ $(i = 1, 2, 3)$, 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 $u \in \mathrm{Pre}(\beta_i)$, azaz $uv = \beta_i$ $(i = 1, 2, 3)$.

Házi feladatok megoldása

1. feladat

Melyik nyelvet generálja a következő nyelvtan? $T = \{(,)\}$
a. $S \to (S) \mid SS \mid \varepsilon$ **b.** $S \to XS \mid \varepsilon$ *és* $X \to (S)$ **c.** $S \to (SS \mid)$

Ha $u \in \mathrm{Pre}(\alpha_1)$, akkor az indukció alapján $\ell_((u) \geq \ell_)(u)$.

Ha $v \in \mathrm{Suf}(\alpha_2)$, akkor $\exists u',\ u'v = \alpha$. Az indukció alapján $\ell_((u') \geq \ell_)(u')$ 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 $\ell_((u) \geq \ell_)(u)$.

Ha $u = \alpha_1\gamma_1$ és $v = \gamma_2\alpha_2$, (azaz γ_1 prefixe valamelyik levezetési szabály jobboldalának) akkor indukció alapján $\ell_((\alpha_1) \geq \ell_)(\alpha_1)$ és könnyen ellenőrizhető, hogy $\ell_((\gamma_1) \geq \ell_)(\gamma_1)$, tehát $\ell_((u) = \ell_((\alpha_1) + \ell_((\gamma_1) \geq \ell_)(\alpha_1) + \ell_)(\gamma_1) = \ell_)(u)$.

$\ell_((\beta_i) = \ell_)(\beta_i)$, 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. $S \to (S)|SS|\varepsilon$ **b.** $S \to XS|\varepsilon$ *és* $X \to (S)$ **c.** $S \to (SS|)$

"$L \supseteq$ **HE**" :
A zárójelek számára vonatkozó indukcóval belátjuk, hogy minden helyes zárójelezés levezethető.
Az üres zárójelezés az $S \to \varepsilon$ 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 $S \to (S) \xrightarrow{*} (w_1)$, a másodikban az $S \to SS \xrightarrow{*} w_1 S \xrightarrow{*} w_1 w_2$ 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!*

$$
\begin{array}{ll}
S' \to \varepsilon \mid S & \\
S \to tSX_t \mid tY_t & \forall t \in T \\
Y_t X_{t'} \to Y_{t'} t & \forall t, t' \in T \\
tX_{t'} \to X_{t'} t & \forall t, t' \in T \\
Y_t \to t &
\end{array}
$$

Házi feladatok megoldása

3. feladat

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

$$
\begin{array}{l}
S \to \varepsilon \mid A \mid aa \\
A \to B \mid ab \mid abA \\
B \to bC \\
C \to \varepsilon \mid B \mid D \\
D \to ab \mid a \mid abD \mid aD \mid abA \mid aA
\end{array}
$$

Normálformák

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

Normálformák (folyt.)

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

0. típusú nyelvtanok

ε-mentesítés

$G = \langle\{a, b\}, \{S, X, Y\}, \mathcal{P}, S\rangle$
$S \to aXSbY \mid ab \mid \varepsilon$
$Xb \to \varepsilon$
$Xa \to aaX$
Végezzük el a 0. típusú ε-mentesítést!

Megoldás:

$S' \to \varepsilon \mid S$
$S \to aXSbY \mid ab$

$aS \to a$	$Sa \to a$	$aXb \to a$	$Xba \to a$
$bS \to b$	$Sb \to b$	$bXb \to b$	$Xbb \to b$
$XS \to X$	$SX \to X$	$XXb \to X$	$XbX \to X$
$YS \to Y$	$SY \to Y$	$YXb \to Y$	$XbY \to Y$
$SS \to S$		$SXb \to S$	$XbS \to S$

$Xa \to aaX$

Kuroda normálforma (1. típusú nyelvtan)

A normálformára alakítás lépései

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

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

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

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

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

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

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

Minden szabályt 3, már helyes újjal helyettesítünk.

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

$AB \to AW$ $\quad AW \to DW$ $\quad DW \to DZ_1$

Chomsky normálforma (2. típus)

0. lépés: Redukció – Zsákutcamentesítés

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

$S \to AB \mid CD$	$C \to Cccc \mid \varepsilon$
$A \to aAa \mid C$	$D \to DD \mid DS$
$B \to bBb \mid C$	$E \to aS \mid bS$

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 $H_1, H_2, \ldots$ halmazsorozat a tartalmazásra nézve monoton növő, felülről korlátos halmazsorozat, így mindig stabilizálódik. Ha $H_{i+1} = H_i$, akkor legyen $H = H_i$.

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

Chomsky normálforma (2. típus)

0. lépés: Redukció – Zsákutcamentesítés (folyt.)

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

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

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

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

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 bejöhetnek. A $K_0, K_1, \ldots$ halmazsorozat a tartalmazásra nézve monoton növő, felülről korlátos halmazsorozat, így mindig stabilizálódik.

$K_{i+1} = K_i \cup \{X \in N \mid \exists Y \to q \in \mathcal{P} : X \subseteq q\}$, 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.)

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

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

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: ε-mentesítés

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

ε-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 ε. 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 ε.

Chomsky normálforma (2. típus)

1. lépés: ε-mentesítés (folyt.)

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

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 ε-szabályokat elhagyjuk.

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

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

$S' \to S \mid \varepsilon$ (ekkor S' lesz az új kezdőszimbólum)

Chomsky normálforma (2. típus)

2. lépés: Láncmentesítés

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

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.)

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

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 jelből áll. (Azaz, a láncszabályokat elhagyjuk.)

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

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

Chomsky normálforma (2. típus)

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

Álterminálisok bevezetése:

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

$S' \to \varepsilon \mid AB \mid Q_aAQ_a \mid Q_aQ_a \mid Q_bBQ_b \mid Q_bQ_b \mid CQ_cQ_cQ_c \mid Q_cQ_cQ_c$
$S \to AB \mid Q_aAQ_a \mid Q_aQ_a \mid Q_bBQ_b \mid Q_bQ_b \mid CQ_cQ_cQ_c \mid Q_cQ_cQ_c$
$A \to Q_aAQ_a \mid Q_aQ_a \mid CQ_cQ_cQ_c \mid Q_cQ_cQ_c$
$B \to Q_bBQ_b \mid Q_bQ_b \mid CQ_cQ_cQ_c \mid Q_cQ_cQ_c$
$C \to CQ_cQ_cQ_c \mid Q_cQ_cQ_c$
$Q_a \to a$ $Q_b \to b$ $Q_c \to c$

Hosszredukció:

Mint a Kuroda NF-nél. Például: $C \to CQ_cQ_cQ_c$ szabály helyett:

$C \to CZ_1$ $Z_1 \to Q_cZ_2$ $Z_2 \to Q_cQ_c$

3. típusú nyelvtanok normálformája

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

1. lépés: **Láncmentesítés**:

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

2. lépés: **Hosszredukció**:

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

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

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