

Formális nyelvek és fordítóprogramok

2. Nyelvek és nyelvtanok II.

Bodó Zalán

Babeş–Bolyai Tudományegyetem
Matematika és Informatika Kar
Magyar Matematika és Informatika Intézet



A Chomsky-féle nyelvosztályok

Átnevezések kiküszöbölése

Normálalakú nyelvtanok

Kiterjesztett nyelvtanok

1-típus

2-típus

3-típus

A Chomsky-féle nyelvosztályok zártsági tulajdonságai

Egyesítés

Szorzat

Iteráció

A Chomsky-féle nyelvosztályok

$$G = (N, T, P, S)$$

Def. Általános/mondatszerkezetű vagy 0-típusú grammatika

– ha nem teszünk semmiféle megkötést a szabályokra vonatkozóan.

Def. Környezetfüggő vagy 1-típusú grammatika

– ha minden szabály $\alpha A \gamma \rightarrow \alpha \beta \gamma$ alakú, ahol

$$A \in N, \alpha, \gamma \in (N \cup T)^*, \beta \in (N \cup T)^+.$$

Megengendhető az $S \rightarrow \varepsilon$ szabály is, ha S nem szerepel egyetlen szabály jobb oldalán sem.

Def. Környezetfüggetlen vagy 2-típusú grammatika

– ha minden szabály $A \rightarrow \beta$ alakú, ahol

$$A \in N, \beta \in (N \cup T)^+.$$

Megengendhető az $S \rightarrow \varepsilon$ szabály is, ha S nem szerepel egyetlen szabály jobb oldalán sem.

A Chomsky-féle nyelvosztályok

Átvezetések kiküszöbölése
Normálalakú nyelvtanok

Kiterjesztett nyelvtanok

1-típus

2-típus

3-típus

A Chomsky-féle nyelvosztályok zártági tulajdonságai

Egyesítés

Szorzat

Iteráció

A Chomsky-féle nyelvosztályok

Átnevezések kiküszöbölése

Normálalakú nyelvtanok

Kiterjesztett nyelvtanok

1-típus

2-típus

3-típus

A Chomsky-féle nyelvosztályok zártági tulajdonságai

Egyesítés

Szorzat

Iteráció

Def. Reguláris vagy 3-típusú grammatika

– ha minden szabály $A \rightarrow aB$ vagy $A \rightarrow a$ alakú, ahol $a \in T$, $A, B \in N$.

Megengendhető az $S \rightarrow \varepsilon$ szabály is, ha S nem szerepel egyetlen szabály jobb oldalán sem.

Jelöljük az előbb definiált nyelvosztályokat \mathcal{L}_i -vel, $i \in \{0, 1, 2, 3\}$.

Tétel

$$\mathcal{L}_0 \supset \mathcal{L}_1 \supset \mathcal{L}_2 \supset \mathcal{L}_3$$

A Chomsky-féle nyelvosztályok

Átnevezések kiküszöbölése

Normálalakú nyelvtanok

Kiterjesztett nyelvtanok

1-típus

2-típus

3-típus

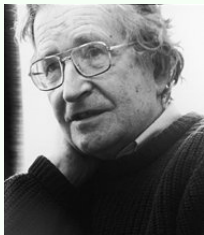
A Chomsky-féle nyelvosztályok zártági tulajdonságai

Egyesítés

Szorzat

Iteráció

Noam Chomsky



Avram Noam Chomsky (1928–) amerikai nyelvész, filozófus, politikai aktivista. A modern nyelvészet és a **generatív grammatika** megteremtője.

Magyarul beszerezhető – és leghíresebb – írásai:

Noam Chomsky. *Mondattani szerkezetek. Nyelv és elme.* Osiris, Budapest, 1999.

https://hu.wikipedia.org/wiki/Noam_Chomsky

https://en.wikipedia.org/wiki/Noam_Chomsky

1. Példa – Környezetfüggő grammatika

$$G_1 = (N_1, T_1, P_1, S_1), N_1 = \{S_1\}, T_1 = \{a, b\}, P_1:$$

$$S_1 \rightarrow aS_1a|a$$

$$S_1aa \rightarrow bS_1aa$$

2. Példa – Környezetfüggetlen grammatika

$$G_2 = (N_2, T_2, P_2, S_2), N_2 = \{S_2, B\}, T_2 = \{0, 1, +, -\}, P_2:$$

$$S_2 \rightarrow B|S_2 + S_2|S_2 - S_2$$

$$B \rightarrow 0|1$$

3. Példa – Reguláris grammatika

$$G_3 = (N_3, T_3, P_3, S_3), N_3 = \{S_3, A\}, T_3 = \{a, b\}, P_3:$$

$$S_3 \rightarrow aS_3|a|bA|b$$

$$A \rightarrow bA|b$$

A Chomsky-féle
nyelvostályok

Átnevezések kiküszöbölése

Normálalakú nyelvtanok

Kiterjesztett
nyelvtanok

1-típus

2-típus

3-típus

A Chomsky-féle
nyelvostályok zártági
tulajdonságai

Egyesítés

Szorzat

Iteráció

Tétel

Minden véges nyelv *reguláris*.

BIZONYÍTÁS. Véges \Rightarrow felsorolhatók a nyelv szavai. Ezért minden $a_1 a_2 \dots a_n$ szóhoz bevezetünk $n - 1$ db. új nemterminálist: A_1, A_2, \dots, A_{n-1} és a köv. szabályokat:

$$S \rightarrow a_1 A_1$$

$$A_1 \rightarrow a_2 A_2$$

...

$$A_{n-1} \rightarrow a_n$$

□

Def. Átnevezés

Az $A \rightarrow B$, $A, B \in N$, alakú szabályt átnevezésnek nevezünk.

ALG 1 Átnevezések kiküszöbölése

- 1: Minden $A \rightarrow B$ és $B \rightarrow C$ esetén vegyük be a szabályok közé az $A \rightarrow C$ átnevezést is. Végezzük ezt addig, amíg P bővíthető.
 - 2: Minden $A \rightarrow B$ és $B \rightarrow \alpha$, $\alpha \notin N$ esetén vegyük fel az $A \rightarrow \alpha$ szabályt is.
 - 3: Töröljük P -ből az átnevezéseket.
-

4. Példa

$G = (\{S, A, B, C\}, \{a, b\}, P, S), P :$

$S \rightarrow A|B$

$A \rightarrow B|D|aB|b$

$B \rightarrow C$

$C \rightarrow B|Aa$

$D \rightarrow C$

Az átnevezések kiküszöbölése után a köv. szabályokat kapjuk:

$S \rightarrow b|aB|Aa$

$A \rightarrow b|aB|Aa$

$B \rightarrow Aa$

Normálalakú nyelvtanok

Def. Normálalakú nyelvtan

Egy grammatika *normálalakú*, ha szabályainak bal oldalán csak nemterminális szimbólumok szerepelnek.

ALG 2 Normálalakba való alakítás

Legyen $G = (N, T, P, S)$, és legyen $T_1 = \{a_1, \dots, a_k\}$ azon terminálisok halmaza, melyek szerepelnek szabályok bal oldalán. Legyenek ekkor $N_1 = \{A_1, \dots, A_k\}$ az új nemterminálisok:

$G' = \{N', T, P', S\}$, ahol

$N' = N \cup N_1$.

Használjuk a következő függvényt: $h : N \cup T \rightarrow N' \cup (T \setminus T_1)$:

$h(a_i) = A_i, a_i \in T_1$

$h(X) = X, X \in N \cup (T \setminus T_1)$. Ekkor

$P' = \{h(\alpha) \rightarrow h(\beta) \mid (\alpha \rightarrow \beta) \in P\} \cup \{A_i \rightarrow a_i \mid i = 1, 2, \dots, k\}$

► **Miért** kell a h leképezést a jobb oldalra is alkalmazni?

5. Példa

$$G = (\{S\}, \{a, b\}, P, S), P: \begin{array}{l} S \rightarrow aSa|a \\ Saa \rightarrow Sb \end{array}$$

Ezt normálalakra hozva a következő grammatikát kapjuk:

$$G' = (\{S, A\}, \{a, b\}, P', S), P':$$

$$\begin{array}{l} S \rightarrow ASA|A \\ SAA \rightarrow Sb \\ A \rightarrow a \end{array}$$

6. Példa

$$G = (\{S, A\}, \{a, b\}, P, S), P: \begin{array}{l} S \rightarrow aAb|bAa \\ aAb \rightarrow aaAbb|ab \\ bAa \rightarrow bbAaa|ba \end{array}$$

Ezt normálalakra hozva a következő grammatikát kapjuk:

$$G' = (\{S, A, C, D\}, \{a, b\}, P', S), P':$$

$$\begin{array}{l} S \rightarrow CAD|DAC \\ CAD \rightarrow CCADD|CD \\ DAC \rightarrow DDACC|DC \\ C \rightarrow a \\ D \rightarrow b \end{array}$$

Kiterjesztett nyelvtanok

Def. 1-típusú kiterjesztett nyelvtan

$\alpha \rightarrow \beta$, ahol $|\alpha| \leq |\beta|$, kivéve esetleg az $S \rightarrow \varepsilon$ szabályt

Def. 2-típusú kiterjesztett nyelvtan

$A \rightarrow \beta$, ahol $A \in N, \beta \in (N \cup T)^*$

Def. 3-típusú kiterjesztett nyelvtan

$A \rightarrow uB$ vagy $A \rightarrow u$, ahol $A, B \in N, u \in T^*$

Tétel.

Minden kiterjesztett nyelvtanhoz megadható egy vele ekvivalens Chomsky-féle, ugyanolyan típusú nyelvtan.

1-típus

$$\alpha \rightarrow \beta, |\alpha| \leq |\beta| \Rightarrow \gamma_1 \delta \gamma_2 \rightarrow \gamma_1 \gamma \gamma_2$$

ALG 3 1-típus

Ha $X_1 X_2 \dots X_m \rightarrow Y_1 Y_2 \dots Y_n, m \leq n$, akkor bevezetjük a köv. nemterminálisokat, A_1, A_2, \dots, A_m és a köv. szabályokat:

$$X_1 X_2 \dots X_m \rightarrow A_1 X_2 \dots X_m$$

$$A_1 X_2 \dots X_m \rightarrow A_1 A_2 \dots X_m$$

...

$$A_1 A_2 \dots A_{m-1} X_m \rightarrow A_1 A_2 \dots A_{m-1} A_m$$

$$A_1 A_2 \dots A_{m-1} A_m \rightarrow Y_1 A_2 \dots A_{m-1} A_m$$

$$Y_1 A_2 \dots A_{m-1} A_m \rightarrow Y_1 Y_2 \dots A_{m-1} A_m$$

...

$$Y_1 Y_2 \dots Y_{m-2} A_{m-1} A_m \rightarrow Y_1 Y_2 \dots Y_{m-2} Y_{m-1} A_m$$

$$Y_1 Y_2 \dots Y_{m-1} A_m \rightarrow Y_1 Y_2 \dots Y_{m-1} Y_m Y_{m+1} \dots Y_n$$

- ▶ Az algoritmus alkalmazása előtt a grammatikát normálalakra kell hozni. (**Miért?**)

A Chomsky-féle nyelvosztályok

Átnevezések kiküszöbölése

Normálalakú nyelvtanok

Kiterjesztett nyelvtanok

1-típus

2-típus

3-típus

A Chomsky-féle nyelvosztályok zártági tulajdonságai

Egyesítés

Szorzat

Iteráció

2-típus

Kiküszöbölni: $A \rightarrow \varepsilon$; esetleg csak $S \rightarrow \varepsilon$ maradhat, ha S nem szerepel szabály jobb oldalán.

U_i = azon nemterminálisok, melyekből levezethető ε

ALG 4 2-típus

- 1: $U_0 = \{A \in N \mid (A \rightarrow \varepsilon) \in P\}$
- 2: $i = 1$
- 3: **repeat**
- 4: $U_i = U_{i-1} \cup \{A \in N \mid (A \rightarrow w) \in P, w \in U_{i-1}^+\}$
- 5: $i = i + 1$
- 6: **until** $U_i = U_{i-1}$

- ▶ $\forall A \rightarrow \alpha, \alpha \neq \varepsilon$ esetén:
 - ▶ vegyük fel az új szabályok közé $(A \rightarrow \alpha)$ -t
 - ▶ vegyük fel az új szabályok közé az összes $(A \rightarrow \beta)$ -t, ahol β -t úgy kaptuk, hogy elhagytunk α -ból 1 v. több U -beli változót, úgy, hogy a jobb oldal nem lett ε
- ▶ ha $\varepsilon \notin L$, akkor VÉGE.

A Chomsky-féle nyelvosztályok

Átnevezések kiküszöbölése

Normálalakú nyelvtanok

Kiterjesztett nyelvtanok

1-típus

2-típus

3-típus

A Chomsky-féle nyelvosztályok zártági tulajdonságai

Egyesítés

Szorzat

Iteráció

A Chomsky-féle nyelvosztályok

Átnevezések kiküszöbölése

Normálalakú nyelvtanok

Kiterjesztett nyelvtanok

1-típus

2-típus

3-típus

A Chomsky-féle nyelvosztályok zártági tulajdonságai

Egyesítés

Szorzat

Iteráció

- ▶ ha $\varepsilon \in L$, akkor:
 - ▶ ha S nem szerepel egyetlen szabály jobb oldalán sem, akkor felvesszük az $S \rightarrow \varepsilon$ szabályt.
 - ▶ különben $S' \rightarrow S$ és $S' \rightarrow \varepsilon$ szabályokat vesszük fel.

3-típus

ALG 5 3-típus

- 1: Alkalmazzuk a 2-típusú nyelvek esetében használt algoritmust az $A \rightarrow \varepsilon$ szabályok kiküszöbölésére.
 - 2: Ezután kiküszöböljük az átnevezéseket a már ismert algoritmus-sal.
 - 3: **for all** $A \rightarrow a_1 a_2 \dots a_n B$, $B \in N \cup \{\varepsilon\}$ **do**
 - 4: Bevezetjük az új A_1, \dots, A_{n-1} nemterminálisokat.
 - 5: Felvesszük a következő szabályokat: $A \rightarrow a_1 A_1$, $A_1 \rightarrow a_2 A_2$,
 \dots , $A_{n-1} \rightarrow a_n B$.
 - 6: **end for**
-

A Chomsky-féle nyelvosztályok zártági tulajdonságai

Tétel

Az $\mathcal{L}_i, i = 0, 1, 2, 3$ nyelvek osztálya zárt a reguláris műveletekre nézve.

BIZONYÍTÁS. Kiterjesztett nyelvtanok segítségével.

Legyen $G_1 = (N_1, T_1, P_1, S_1)$ és $G_2 = (N_2, T_2, P_2, S_2)$ két i -típusú kiterjesztett grammatika. Feltételezzük, hogy $N_1 \cap N_2 = \emptyset$

Egyesítés

Minden nyelv esetén ugyanaz.

$$G_U = (N_1 \cup N_2 \cup \{S\}, T_1 \cup T_2, P_1 \cup P_2 \cup \{S \rightarrow S_1, S \rightarrow S_2\}, S)$$

$L(G_U) = L(G_1) \cup L(G_2)$ és ha G_1, G_2 i -típusú, akkor G_U is az.

Ha $\varepsilon \in L(G_i)$, akkor töröljük az $S_i \rightarrow \varepsilon$ -t ($i \in \{1, 2\}$) és bevesszük helyette az $S \rightarrow \varepsilon$ szabályt.

A Chomsky-féle
nyelvosztályok

Átnevezések kiküszöbölése

Normálalakú nyelvtanok

Kiterjesztett
nyelvtanok

1-típus

2-típus

3-típus

A Chomsky-féle
nyelvosztályok zártági
tulajdonságai

Egyesítés

Sorozat

Iteráció

Szorzat

$$G_x = (N_1 \cup N_2 \cup \{S\}, T_1 \cup T_2, P_1 \cup P_2 \cup \{S \rightarrow S_1 S_2\}, S)$$

$$L(G_x) = L(G_1)L(G_2)$$

$i = 0, 2$ G_x is ilyen típusú lesz

- $i = 1$
- ▶ ha $(S_1 \rightarrow \varepsilon) \in P_1$, $(S_2 \rightarrow \varepsilon) \notin P_2$, akkor az $S_1 \rightarrow \varepsilon$ szabályt az $S \rightarrow S_2$ szabállyal helyettesítjük.
 - ▶ a szimmetrikus esetben hasonlóan járunk el.
 - ▶ ha P_1 és P_2 -ben is van $S \rightarrow \varepsilon$, akkor ezeket helyettesítjük az $S \rightarrow \varepsilon$ szabállyal.

$i = 3$ $S \rightarrow S_1 S_2$ nem reguláris szabály:

$G_x = (N_1 \cup N_2, T_1 \cup T_2, P'_1 \cup P_2, S_1)$, P'_1 abban P_1 -től, hogy az $A \rightarrow u$, $u \in T^*$ szabályok helyett az $A \rightarrow uS_2$ szabályok kerülnek P'_1 -be.

Iteráció

$$G_* = (N_1 \cup \{S\}, T_1, P, S)$$

$$i = 2 \quad P = P_1 \cup \{S \rightarrow S_1 S, S \rightarrow \varepsilon\}$$

$i = 3 \quad P = P'_1 \cup \{S \rightarrow S_1, S \rightarrow \varepsilon\}$, ahol P'_1 abban különbözik P_1 -től, hogy minden $A \rightarrow u, u \in T^*$ szabály helyett $A \rightarrow uS$ alakút veszünk.

$i = 0, 1$ nem jók az eddigi módszerek, mert megtörténhet a következő:

$S \xrightarrow{*} S_1 S_1 \alpha, S_1 \xrightarrow{*} \alpha_1 \beta_1, S_1 \xrightarrow{*} \alpha_2 \beta_2$, és $\beta_1 \alpha_2$ egy helyettesítési szabály bal oldala; ekkor az $S \xrightarrow{*} \alpha_1 \underbrace{\beta_1 \alpha_2}_{\beta_2 \alpha} \beta_2 \alpha$ levezetésben olyan szót is

generálhatunk, ami nincs benne az iterált nyelvben.

Megoldás:

- ▶ normálalakú nyelvtanra való alakítás
- ▶ $N_* = N_1 \cup \{S, S'\}$
- ▶ $P = P_1 \cup \{S \rightarrow \varepsilon, S \rightarrow S_1 S'\} \cup \{aS' \rightarrow aS \mid a \in T_1\}$

A Chomsky-féle
nyelvosztályok

Átnevezések kiküszöbölése
Normálalakú nyelvtanok

Kiterjesztett
nyelvtanok

1-típus

2-típus

3-típus

A Chomsky-féle
nyelvosztályok zártági
tulajdonságai

Egyesítés

Szorzat

Iteráció

Példa: iteráció

$$\begin{array}{l}
 S \rightarrow ABA \\
 BA \rightarrow AB \\
 A \rightarrow a \\
 B \rightarrow b
 \end{array}
 \Rightarrow L = \{aba, aab\}$$

Iteráció:

$$\begin{array}{l}
 S \xrightarrow{*} SS \Rightarrow ABAABA \Rightarrow AABABA \Rightarrow AAABBA \\
 \Rightarrow AAABAB \Rightarrow AAAABB \Rightarrow aaaabb \notin L^*
 \end{array}$$