

0. Chomsky-nf.
1. ε -mentesítés
2. láncszabály-mentesítés
3. jobboldalak helyreigazítása

0. Chomsky-nf.

Nézzük meg a környezetfüggetlen nyelvtanok Chomsky normálformáját, hogy mi is az.

A G környezetfüggetlen nyelvtan Chomsky normálformában van, ha szabályai $A \rightarrow BC$, $A \rightarrow a$ alakúak, ahol A, B, C nemterminálisok és $a \in \Sigma$ egy terminális szimbólum. Esetleg szerepelhet az $S \rightarrow \varepsilon$ szabály is, de ekkor S kezdőszimbólum és nem szerepel egyetlen szabály jobb oldalán sem.

Minden G környezetfüggetlen nyelvtanhoz megadható egy Chomsky normálformában levő G' környezetfüggetlen nyelvtan, amely ugyanazt a nyelvet generálja, mint G . Hogyan is megy ez?

Sorrendben:

1. G ε -mentesítése;
2. A kapott nyelvtan láncszabálymentesítése;
3. Azon szabályok cseréje, melyek jobb oldalain nem kettő nemterminális vagy egy terminális szerepel. (tekintettel $S \rightarrow \varepsilon$ -ra)

A kapott G' nyelvtan Chomsky normálformában lesz. Az egyes lépések részletesebben:

1. ε -mentesítés

G ε -mentesítése:

Legyen $G = (N, \Sigma, P, S)$ környezetfüggetlen nyelvtan.

- Legyen H azon nemterminálisok halmaza, melyekből ε levezethető.

Azaz legyen $H_0 = \{A \mid A \rightarrow \varepsilon \in P\}$ és $H_{n+1} = H_n \cup \{B \mid B \rightarrow \alpha \in P, \alpha \in H_n^*, \alpha \neq \varepsilon\}$.

Ekkor van olyan k , hogy $H_k = H_{k+1}$. (mivel $H_n \subseteq H_{n+1}$ és a N véges)

Legyen hát $H = H_k$.

- Ezután P -ből elhagyjuk az összes $A \rightarrow \varepsilon$ alakú szabályt és minden $B \rightarrow w$ alakú szabályt helyettesítünk olyan $B \rightarrow w'$ szabályokkal, melyekre igaz, hogy $w' \neq \varepsilon$ úgy áll elő w -ből, hogy valamennyi (esetleg nulla) darab H -beli nemterminálíst törölünk.

- Ha $S \in H$ akkor legyen S' egy új nemterminális és $S' \rightarrow S \mid \varepsilon$ új szabályok. Az új kezdőszimbólum S' . Egyébként marad S a kezdőszimbólum.

A kapott G' nyelvtan ugyanazt a nyelvet generálja mint G és nincsenek benne $A \rightarrow \varepsilon$ alakú szabályok, legfeljebb $S' \rightarrow \varepsilon$. De ekkor S' nem fordul elő egyetlen szabály jobb oldalán sem (hiszen új szimbólum).

Példa: legyen G a következő szabályokkal adott:

$$S \rightarrow BA | AbS | CC$$

$$A \rightarrow BB | bA$$

$$B \rightarrow \varepsilon | ab$$

$$C \rightarrow aB$$

$$H_1 = \{B\}, H_2 = \{B, A\}, H_3 = \{B, A, S\} = H_4. \text{ Így } H = \{B, A, S\}.$$

Az ε -mentes G' nyelvtan:

$$S' \rightarrow S | \varepsilon$$

$$S \rightarrow BA | B | A | AbS | Ab | bS | b | CC$$

$$A \rightarrow BB | B | bA | b$$

$$B \rightarrow ab$$

$$C \rightarrow aB | a$$

Itt S' a kezdőszimbólum.

Ekkor $L(G) = L(G')$.

2. láncszabálymentesítés

Legyen $G = (N, \Sigma, P, S)$ környezetfüggetlen nyelvtan. Határozzuk meg minden $A \in N$ -re azon nemterminálisok halmazát, melyek levezethetők A -ból. Azaz legyen

$$N_0^A = \{A\} \cup \{B \in N \mid A \rightarrow B \in P\} \text{ és } N_{n+1}^A = N_n^A \cup \{C \in N \mid B \rightarrow C \in P, \text{ valamely } B \in N_n^A\text{-re}\}.$$

Ekkor van olyan k , hogy $N_k^A = N_{k+1}^A$. (mivel N véges és $N_n^A \subseteq N_{n+1}^A$) Legyen hát $N^A = N_k^A$.

Ekkor N^A azon nemterminálisok halmaza, melyek levezethetők A -ból.

Most hagyjuk el az összes $A \rightarrow B$ alakú szabályt P -ből és vegyük az összes olyan $A \rightarrow w$, $w \notin N$ szabályt, melyre igaz, hogy van olyan $B \in N^A$, melyre $B \rightarrow w \in P$.

Az így kapott szabályok halmaza legyen P'' .

Ekkor a $G'' = (N, \Sigma, P'', S)$ *láncszabálymentes*, azaz nem tartalmaz $A \rightarrow B$ alakú szabályt, ahol $B \in N$.

Példa: Vegyük az előző példában szereplő ε -mentes G' nyelvtant:

$$\begin{aligned}
S' &\rightarrow S|\varepsilon \\
S &\rightarrow BA|B|A|AbS|Ab|bS|b|CC \\
A &\rightarrow BB|B|bA|b \\
B &\rightarrow ab \\
C &\rightarrow aB|a
\end{aligned}$$

Itt S' a kezdőszimbólum. Láncszabálymentesítsük!

$$\begin{aligned}
N^{S'} &= \{S', S, A, B, C\}, \\
N^S &= \{S, A, B, C\}, \\
N^A &= \{A, B\}, \\
N^B &= \{B\}, \\
N^C &= \{C\}.
\end{aligned}$$

A kapott nyelvtan szabályai:

$$\begin{aligned}
S' &\rightarrow \varepsilon|BA|AbS|Ab|bS|b|CC|BB|bA|b|ab|aB \\
S &\rightarrow BA|AbS|Ab|bS|b|CC|BB|bA|b|ab|aB \\
A &\rightarrow BB|bA|b|ab \\
B &\rightarrow ab \\
C &\rightarrow aB|a
\end{aligned}$$

Itt S' a kezdőszimbólum. Ez a nyelvtan ugyanazt a nyelvet generálja, mint G' , azaz ugyanazt a nyelvet generálja, mint G .

3. A jobb oldalak helyreigazítása

Legyen adott $G = (N, \Sigma, P'', S')$. Vezessünk be minden Σ -beli a betűhöz egy új X_a nemterminálist. Legyen $A \rightarrow w \in P''$. Ha w hossza legalább kettő és szerepel benne egy a terminális szimbólum, akkor cseréljük le az összes előfordulását X_a -ra. Továbbá vegyük fel az $X_a \rightarrow a$ szabályokat.

Példa: Az előbbi példában szereplő ε és láncszabálymentes nyelvtan esetén:

$$\begin{aligned}
S' &\rightarrow \varepsilon|BA|AX_bS|AX_b|X_bS|b|CC|BB|X_bA|b|X_aX_b|X_aB \\
S &\rightarrow BA|AX_bS|AX_b|X_bS|b|CC|BB|X_bA|b|X_aX_b|X_aB \\
A &\rightarrow BB|X_bA|b|X_aX_b \\
B &\rightarrow X_aX_b \\
C &\rightarrow X_aB|a \\
X_a &\rightarrow a \\
X_b &\rightarrow b
\end{aligned}$$

Végül, a túl hosszú jobb oldalakat feldaraboljuk új nemterminálisok bevezetésével.

Pontosabban ha $A \rightarrow A_1 A_2 \dots A_n$ egy szabály, ahol $n \geq 3$, akkor cseréljük le a következő szabályokra: $A \rightarrow A_1 Y_1$, $Y_1 \rightarrow A_2 Y_2$, ..., $Y_{n-3} \rightarrow A_{n-2} Y_{n-2}$, $Y_{n-2} \rightarrow A_{n-1} A_n$.

Példa:

$$\begin{aligned} S' &\rightarrow \varepsilon | BA | AY_1 | AX_b | X_b S | b | CC | BB | X_b A | b | X_a X_b | X_a B \\ S &\rightarrow BA | AY_1 | AX_b | X_b S | b | CC | BB | X_b A | b | X_a X_b | X_a B \\ A &\rightarrow BB | X_b A | b | X_a X_b \\ B &\rightarrow X_a X_b \\ C &\rightarrow X_a B | a \\ X_a &\rightarrow a \\ X_b &\rightarrow b \\ Y_1 &\rightarrow X_b S \end{aligned}$$

Az így kapott nyelvtan ugyanazt a nyelvet generálja, mint G . Az algoritmusnak itt vége.

Ha adott egy környezetfüggetlen nyelvtan és sorrendben elvégezzük rajta az ε , majd a láncszabálymentesítést és végül a jobb oldalakat helyreigazítjuk, akkor az így kapott nyelvtan szabályainak jobb oldalain mi is állhat?

Ugye nem lehet egy darab nemterminális, hiszen láncszabálymentesítettünk. Ha ε , akkor az adott szabály bal oldalán a kezdőszimbólum áll, mely nem szerepel egy szabály jobb oldalán sem. Továbbá nem állhat 2-nél hosszabb szó, mert feldaraboltuk a jobb oldalakat. Az sem lehet igaz, hogy egy olyan szó szerepel ott, melyben van terminális és nemterminális szimbólum is, hiszen új X_a nemterminálisokra cseréltük a terminálisokat.

A szabályok jobb oldalain vagy BC áll, ahol B, C valami nemterminálisok, vagy a áll, ahol a terminális. Vagy ε áll, de ekkor a bal oldalon a kezdőszimbólum szerepel, mely nem fordul elő egy szabály jobb oldalán sem. A kapott nyelvtan Chomsky normálformában van.