

Hogyan készítsünk Chomsky normál alakú (CNF) nyelvtant környezetfüggetlenből?

Legelőször tennék a címhez még egy kiegészítést. Ugyanis jól fésült nyelvtant fogunk csak átalakítani. Ha a nyelvtanunk nem jól fésült, vagyis tartalmaz egyszeres szabályt, ε -szabályt, vagy felesleges szimbólumot, akkor a már jól ismert fésülés eljárásainkat kell először alkalmaznunk.

Definíció

A nyelvtan Chomsky normál alakú (CNF), ha minden szabálya

I. $A \rightarrow BC$, illetve

II. $A \rightarrow a$ alakú.

(A, B, C nemterminális, a pedig terminális szimbólum.) A nyelvtanban az $S \rightarrow \varepsilon$ szabály is benne lehet (S a mondat-szimbólumot jelöli), amennyiben a generált nyelv tartalmazza az üres jelsorozatot. Ekkor viszont S nem szerepelhet szabály jobb oldalán.

Tétel

Minden jól fésült környezetfüggetlen nyelvtan Chomsky normál alakúvá alakítható.

Bizonyítás

A bizonyítás konstruktív lesz, azaz egy jól fésült CF nyelvtant transzformálunk Chomsky normál alakúra.

1. Az olyan szabályokkal, amelyek már kezdetben eleget tesznek a formai kívánalmaknak, a továbbiakban nem foglalkozunk; ezeket nem kell átalakítani.
2. Minden terminális szimbólum mellé vegyünk fel egy nemterminálist! (Az a -nak megfelelő nemterminálist jelöljük $\hat{A}$ -val.)
 - Az összes helyen, ahol az adott terminális szerepel szabály jobb oldalán, cseréljük le a neki megfelelő nemterminálissra.
 - A szabályhalmazhoz adjunk hozzá annyi szabályt, ahány újonnan bevezetett nemterminális szimbólum van. Az új szabályok formája $\hat{A} \rightarrow a$ legyen.

Triviális, hogy a változtatott nyelvtan által generált nyelv nem változott, mivel most két lépésben gyártunk terminális szimbólumokat az eddigi egy lépés helyett.

3. Most már csak olyan szabályaink lehetnek, melyek $B \rightarrow B_1 B_2 \dots B_n$ alakúak. (A B_i -k nemterminális szimbólumok). Itt n nem lehet 0, mivel az eredeti nyelvtanunk ε -szabály mentesített (jól fésült) volt, és $n = 1$ sem lehet, mivel a jól fésültség miatt egyszeres szabályaink sem lehettek. Végül elmondhatjuk, hogy $n = 2$ sem lehet, mivel korábban már úgy döntöttünk, hogy a CNF előírásainak megfelelő szabályokkal a továbbiakban nem foglalkozunk. Következésképpen $n > 2$.

Ekkor viszont megcsinálhatjuk a következő átalakítást:

$$B \rightarrow B_1 B_2 \dots B_n \rightsquigarrow \begin{cases} B & \rightarrow B_1 \hat{B}_1 \\ \hat{B}_1 & \rightarrow B_2 \dots B_n \end{cases}$$

ahol $\hat{B}_1$ új, eddig nem szereplő nemterminális.

Az átalakítás utáni két szabály közül a felső már I. típusú, míg az alsó jobboldalának hossza eggyel rövidebb lett, mint korábban volt. Ezt az eljárást még egy párszor $((n-2)$ -szer) alkalmazva, az n előbb-utóbb 2-re lecsökken, és ekkor kész vagyunk. Arra azért kell figyelnünk, hogy mindig új, eddig nem szerepelt nemterminálist hozzunk be. Itt ismét csak triviális, hogy az átalakítás előtt, illetve után ugyanazt a nyelvet tudtuk generálni nyelvtanainkkal.

Az átalakítás szemléltetése egy példán

A feladat legyen a következő nyelvtan CNF-be transzformálása!

$$\begin{aligned} S &\rightarrow aAbBc \\ A &\rightarrow aAb \mid ab \\ B &\rightarrow Bc \mid c \end{aligned}$$

0. Ez jól fésült nyelvtan, tehát nekiláthatunk az előbb leírt eljárásnak.

1. Kiválasztjuk a nem CNF alakú szabályokat:

$$\begin{aligned} S &\rightarrow aAbBc \\ A &\rightarrow aAb \mid ab \\ B &\rightarrow Bc \end{aligned}$$

$B \rightarrow c$ már kezdetben jó alakú (II. típusú) volt.

2. Külön lépésben generáljuk a terminálisokat:

$$\begin{aligned} S &\rightarrow \hat{A}\hat{A}\hat{B}\hat{B}\hat{C} \\ A &\rightarrow \hat{A}\hat{A}\hat{B} \mid \hat{A}\hat{B} \\ B &\rightarrow \hat{B}\hat{C} \mid c \\ \hat{A} &\rightarrow a \\ \hat{B} &\rightarrow b \\ \hat{C} &\rightarrow c \end{aligned}$$

Ekkor már csak két szabály nem felel meg a CNF előírásainak:

$$\begin{aligned} S &\rightarrow \hat{A}\hat{A}\hat{B}\hat{B}\hat{C} \\ A &\rightarrow \hat{A}\hat{A}\hat{B} \end{aligned}$$

A továbbiakban értelemszerűen csak ezekkel foglalkozunk.

3. A fent leírtakat kellően sokszor végrehajtva a két szabály a következőképpen fog átalakulni:

$$S \rightarrow \hat{A}\hat{A}\hat{B}\hat{B}\hat{C} \rightsquigarrow \begin{cases} S & \rightarrow \hat{A}S_1 \\ S_1 & \rightarrow \hat{A}S_2 \\ S_2 & \rightarrow \hat{B}S_3 \\ S_3 & \rightarrow \hat{B}\hat{C} \end{cases}$$

$$A \rightarrow \hat{A}A\hat{B} \rightsquigarrow \begin{cases} A \rightarrow \hat{A}A_1 \\ A_1 \rightarrow A\hat{B} \end{cases}$$

A végeredményként kapott szabályhalmaz:

$$\begin{aligned} S &\rightarrow \hat{A}S_1 \\ S_1 &\rightarrow AS_2 \\ S_2 &\rightarrow \hat{B}S_3 \\ S_3 &\rightarrow B\hat{C} \\ A &\rightarrow \hat{A}A_1 \mid \hat{A}\hat{B} \\ A_1 &\rightarrow A\hat{B} \\ B &\rightarrow B\hat{C} \mid c \\ \hat{A} &\rightarrow a \\ \hat{B} &\rightarrow b \\ \hat{C} &\rightarrow c \end{aligned}$$