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 fésülős eljárásainkat kell először alkalmaznunk.

Definíció

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

I. $A \rightarrow BC$*, illetve*

II. $A \rightarrow \mathsf{a}$ *alakú.*

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

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 megfeleltetett nemterminálist jelöljük $\hat{A}$-val.)

   - Az összes helyen, ahol az adott terminális szerepel szabály jobboldalán, cseréljük le a neki megfelelő nemterminálisra.
   - A szabályhalmazhoz adjunk hozzá annyi szabályt, ahány újonnan bevezetett nemterminális szimbólum van. Az új szabályok formája $\hat{A} \rightarrow \mathsf{a}$ 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 $B \rightarrow B_1B_2 \ldots B_n$ 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:

$$B \to B_1 B_2 \ldots B_n \quad \rightsquigarrow \quad \left\{ \begin{array}{lcl} B & \to & B_1 \hat{B}_1 \\ \hat{B}_1 & \to & B_2 \ldots B_n \end{array} \right.$$

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

Az átalakítás utáni két szabály közül a fölső 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!

$$\begin{array}{lcl} S & \to & \mathtt{a}A\mathtt{b}B\mathtt{c} \\ A & \to & \mathtt{a}A\mathtt{b} \mid \mathtt{ab} \\ B & \to & B\mathtt{c} \mid \mathtt{c} \end{array}$$

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

1. Kiválasztjuk a nem CNF alakú szabályokat:

$$\begin{array}{lcl} S & \to & \mathtt{a}A\mathtt{b}B\mathtt{c} \\ A & \to & \mathtt{a}A\mathtt{b} \mid \mathtt{ab} \\ B & \to & B\mathtt{c} \end{array}$$

   $B \to \mathtt{c}$ már kezdetben jó alakú (II. típusú) volt.

2. Külön lépésben generáljuk a terminálisokat:

$$\begin{array}{lcl} S & \to & \hat{A}A\hat{B}B\hat{C} \\ A & \to & \hat{A}A\hat{B} \mid \hat{A}\hat{B} \\ B & \to & B\hat{C} \mid \mathtt{c} \\ \hat{A} & \to & \mathtt{a} \\ \hat{B} & \to & \mathtt{b} \\ \hat{C} & \to & \mathtt{c} \end{array}$$

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

$$\begin{array}{lcl} S & \to & \hat{A}A\hat{B}B\hat{C} \\ A & \to & \hat{A}A\hat{B} \end{array}$$

   A továbbiakban értelemszerűen csak ezekkel foglalkozunk.

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

$$S \to \hat{A}A\hat{B}B\hat{C} \quad \rightsquigarrow \quad \left\{ \begin{array}{lcl} S & \to & \hat{A}S_1 \\ S_1 & \to & AS_2 \\ S_2 & \to & \hat{B}S_3 \\ S_3 & \to & B\hat{C} \end{array} \right.$$

$$A \to \hat{A}A\hat{B} \quad \rightsquigarrow \quad \left\{ \begin{array}{lcl} A & \to & \hat{A}A_1 \\ A_1 & \to & A\hat{B} \end{array} \right.$$

A végeredményként kapott szabályhalmaz:

$$\begin{array}{lcl} S & \to & \hat{A}S_1 \\ S_1 & \to & AS_2 \\ S_2 & \to & \hat{B}S_3 \\ S_3 & \to & B\hat{C} \\ A & \to & \hat{A}A_1 \mid \hat{A}\hat{B} \\ A_1 & \to & A\hat{B} \\ B & \to & B\hat{C} \mid \mathtt{c} \\ \hat{A} & \to & \mathtt{a} \\ \hat{B} & \to & \mathtt{b} \\ \hat{C} & \to & \mathtt{c} \end{array}$$