Bezkontextové jazyky 2/3

Transformace bezkontextových gramatik

Ekvivalentní gramatiky

Definice 6.1 Nechť G_1 a G_2 jsou gramatiky libovolného typu Chomského klasifikace. G_1 a G_2 jsou ekvivalentní, pokud $L(G_1) = L(G_2)$.

Věta 6.1 Nechť $G = (N, \Sigma, P, S)$ je gramatika, $A \to \alpha B \beta$, $B \in N$, $\alpha, \beta \in (N \cup \Sigma)^*$ je pravidlo z P a nechť $B \to \gamma_1 \mid \gamma_2 \mid \ldots \mid \gamma_n$ jsou všechna B-pravidla z P. Pak gramatika $G' = (N, \Sigma, P', S)$ kde

$$P' = P \setminus \{A \to \alpha B \beta\} \cup \{A \to \alpha\gamma_1\beta, A \to \alpha\gamma_2\beta, \ldots, A \to \alpha\gamma_n\beta\}$$

je ekvivalentní s gramatikou G.

Důkaz. Na cvičení. $\square$

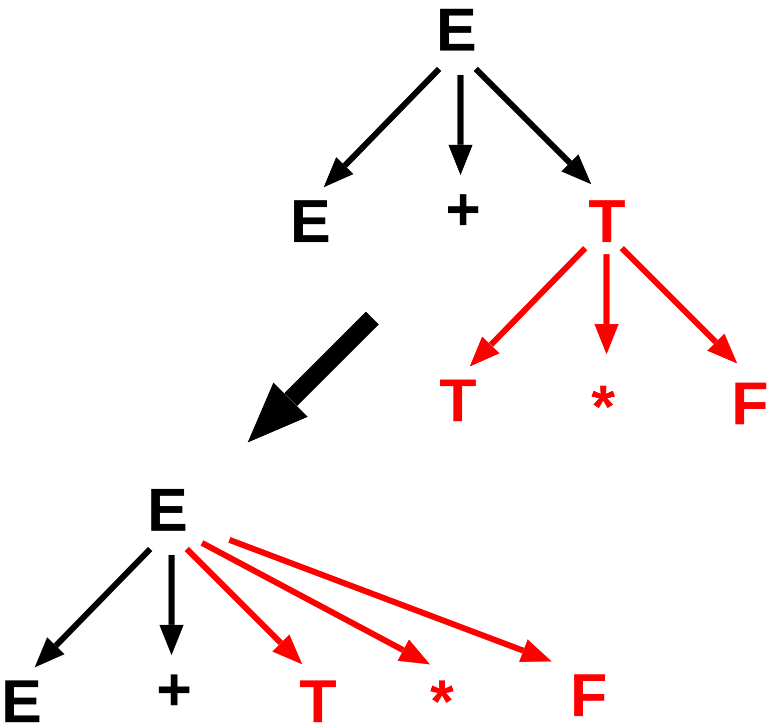
Příklad 6.1
Gramatiky s pravidly

$E \to E + T \mid T$
$T \to T * F \mid F$
$F \to (E) \mid i$

resp.

$E \to E + T * F \mid E + F \mid T$
$T \to T * F \mid F$
$F \to (E) \mid i$

jsou ekvivalentní.



Nedostupné a zbytečné symboly

Definice 6.2 Nechť $G = (N, \Sigma, P, S)$ je gramatika a $X \in N \cup \Sigma$ symbol. Říkáme, že symbol X je nedostupný v G, jestliže v G neexistuje derivace $S \Rightarrow^* \alpha X \beta$ pro nějaké $\alpha, \beta \in (N \cup \Sigma)^*$. Symbol X nazýváme zbytečný v G, jestliže v G neexistuje derivace tvaru $S \Rightarrow^* \alpha X \beta \Rightarrow^* zxy$ pro nějaké $\alpha, \beta \in (N \cup \Sigma)^*$ a $zxy \in \Sigma^*$.

Příklad 6.2 Uvažujme gramatiku $G = (\{S, A, B\}, \{a, b\}, P, S)$ s pravidly:

$S \rightarrow SB \mid a$
$A \rightarrow b$
$B \rightarrow Ba$

Symboly A, B, b jsou zbytečné. Symboly A, b jsou nedostupné.

Poznámka 6.1 $G' = (\{S\}, \{a\}, \{S \rightarrow a\}, S)$ je ekvivalentní s G.

Nonterminály generující terminální řetězce

Algoritmus 6.1 Výpočet množiny nonterminálů generujících terminální řetězce
Vstup: Gramatika $G = (N, \Sigma, P, S)$.
Výstup: Množina $N_t = \{A \mid A \Rightarrow^+ w, w \in \Sigma^*\}$.
Metoda: Počítáme množiny $N_0, N_1, N_2, \ldots$ rekurentně takto:

1. N_0 := $\emptyset, i = 1$
2. N_i := $\{A \mid A \rightarrow \alpha$ je v P a $\alpha \in (N_{i-1} \cup \Sigma)^*\}$
3. Je-li $N_i \neq N_{i-1}$, i := $i + 1$ a vrať se k (2). Je-li $N_i = N_{i-1}$, polož $N_t = N_i$ a skonči.

Dostupné symboly

Algoritmus 6.2 Výpočet množiny dostupných symbolů
Vstup: Gramatika $G = (N, \Sigma, P, S)$.
Výstup: Množina $V = \{X \mid S \Rightarrow^* \alpha X \beta, \alpha, \beta \in (N \cup \Sigma)^*\}$.
Metoda:

1. V_0 := $\{S\}, i = 1$
2. V_i := $\{X \mid A \rightarrow \alpha X \beta$ je v P a $A \in V_{i-1}\} \cup V_{i-1}$
3. Je-li $V_i \neq V_{i-1}$, i := $i + 1$ a vrať se k (2). Je-li $V_i = V_{i-1}$, polož $V = V_i$ a skonči.

Odstranění zbytečných symbolů

Algoritmus 6.3 Odstranění zbytečných symbolů
Vstup: Gramatika $G = (N, \Sigma, P, S)$.
Výstup: Gramatika $G' = (N', \Sigma', P', S)$ bez zbytečných symbolů, $L(G) = L(G')$.
Metoda:

1. Aplikací algoritmu 4.1 na G vypočti množinu N_t.
2. Polož $\overline{G} = (N_t \cup \{S\}, \Sigma, \overline{P}, S)$, kde
   $\overline{P} = \{A \rightarrow \alpha | (A \rightarrow \alpha) \in P \wedge A \in N_t \wedge \alpha \in (N_t \cup \Sigma)^*\}$.
3. Aplikací algoritmu 4.2 na $\overline{G}$ vypočti množinu V.
4. Výslednou gramatiku G' sestroj takto:
   (a) $N' = N_t \cap V$
   (b) $\Sigma' = \Sigma \cap V$
   (c) $P' = \{A \rightarrow \alpha | (A \rightarrow \alpha) \in P \wedge A \in N' \wedge \alpha \in V^*\}$

Poznámka: Sjednocení $N_t \cup \{S\}$ v bodě 2 řeší případ, kdy $L(G) = \emptyset$ a $N_t = \emptyset$, ovšem gramatika musí mít svůj startovací symbol.

Příklad 6.3 Uvažujme gramatiku
$G = (\{S, A, B\}, \{a, b\}, \{S \to a, S \to A, A \to AB, B \to b\}, S)$.

1. $N_0 = \emptyset$, $N_1 = \{S, B\}$, $N_2 = N_1 = N_t = \{S, B\}$
2. $\overline{G} = (\{S, B\}, \{a, b\}, \{S \to a, B \to b\}, S)$
3. $V_0 = \{S\}$, $V_1 = \{S, a\}$, $V_2 = V_1 = V = \{S, a\}$
4. $G' = (\{S\}, \{a\}, \{S \to a\}, S)$.

Poznámka 6.2 Pořadí kroků 2. a 4. je významné.

Odstranění ε-pravidel

Definice 6.3 G je gramatikou bez ε-pravidel, jestliže neobsahuje žádné ε-pravidlo (pravidlo tvaru $A \rightarrow \varepsilon$), nebo, pokud $\varepsilon \in L(G)$, potom obsahuje jediné ε-pravidlo $S \rightarrow \varepsilon$ a S se pak nevyskytuje na pravé straně žádného přepisovacího pravidla.

Algoritmus 6.4 Transformace na gramatiku bez ε-pravidel
Vstup: Gramatika $G = (N, \Sigma, P, S)$.
Výstup: Gramatika $G' = (N', \Sigma', P', S')$ bez ε-pravidel ekvivalentní s G.
Metoda:

1. Vypočítej množinu $N_\varepsilon = \{A \mid A \Rightarrow^+ \varepsilon\}$
2. Každé pravidlo z P, které není ε-pravidlem, uvažuj ve tvaru $A \rightarrow \alpha_0 B_1 \alpha_1 B_2 \ldots B_k \alpha_k$, kde $B_i \in N_\varepsilon, \alpha_i \in (N \setminus N_\varepsilon \cup \Sigma)^*$ pro $i = 1, \ldots, k$ Toto pravidlo nahraď množinou pravidel, které vzniknou všemi možnými substitucemi $B_i \rightsquigarrow B_i$ a $B_i \rightsquigarrow \varepsilon$ pro $i = 1, \ldots, k$ (to jest substitucemi, kdy nonterminály z N_ε jsou alternativně ponechávány a vypouštěny). Počet těchto substitucí (nových pravidel) je zřejmě 2^k.
3. Všechna ε-pravidla vypusť.
4. Pokud $S \in N_\varepsilon$, pak $N' = N \cup \{S'\}$, kde S' je nový nonterminální symbol, a přidej pravidla $S' \rightarrow \varepsilon \mid S$, v opačném případě $N' = N$, $S' = S$

Příklad 6.4 Uvažujme gramatiku $G = (\{A, B, C\}, \{a, b, c\}, P, A)$ s pravidly:

$A \to AbAcBC \mid \varepsilon \mid a$
$B \to b \mid \varepsilon$
$C \to c \mid \varepsilon$

1. $N_\varepsilon = \{A, B, C\}$
2. $A \to AbAcBC$
   $A \to \ bAcBC$
   $A \to Ab\ cBC$
   $A \to \ b\ cBC$
   $\vdots$
   $A \to \ bc$
   $A \to a$
   $B \to b$
   $C \to c$
3. $A' \to \varepsilon$
   $A' \to A$

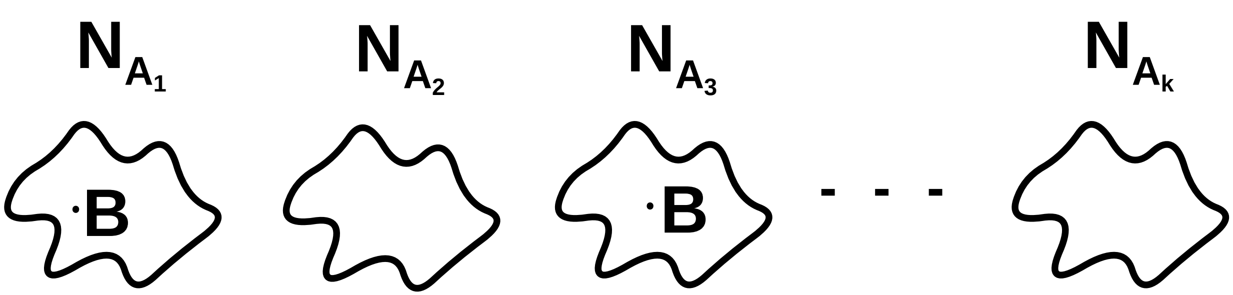
A' je nový startovací symbol.

Odstranění jednoduchých pravidel

Definice 6.4 Pravidlo tvaru $A \rightarrow B$, kde $A, B \in N$ nazýváme jednoduché pravidlo.

Algoritmus 6.5 Transformace na gramatiku bez jednoduchých pravidel
Vstup: Gramatika $G = (N, \Sigma, P, S)$ bez ε-pravidel.
Výstup: Gramatika $G' = (N, \Sigma, P', S)$ bez jednoduchých pravidel ekvivalentní s G.
Metoda:

1. Pro všechny $A \in N$ vypočítej množinu $N_A = \{B \mid A \Rightarrow^* B\}$, polož $P' := \emptyset$.
2. Nechť $B \rightarrow \alpha$, $\alpha \notin N$ je pravidlo z P. Potom k P' přidej nová pravidla $A_i \rightarrow \alpha$ pro všechny A_i, kde $B \in N_{A_i}$.
3. Výsledná množina pravidel P' tvoří všechna pravidla gramatiky G' (neobsahuje jednoduchá pravidla).



Nová pravidla: $A_1 \rightarrow \alpha$ $\qquad$ $A_3 \rightarrow \alpha$

Příklad 6.5 Uvažujme gramatiku $G = (\{E, T, F\}, \{i, +, *, (,)\}, P, E)$ s pravidly:
$E \rightarrow E + T \mid T$
$T \rightarrow T * F \mid F$
$F \rightarrow (E) \mid i$

1. Nalezneme množiny N_A pro všechny $A \in N$:
   $N_E = \{E, T, F\}$ $N_T = \{T, F\}$ $N_F = \{F\}$
2. Doplňujeme nová pravidla a vypouštíme jednoduchá pravidla:
   $E \rightarrow E + T \mid T * F \mid (E) \mid i$
   $T \rightarrow T * F \mid (E) \mid i$
   $F \rightarrow (E) \mid i$

Cyklus

Definice 6.5 Nechť $G = (N, \Sigma, P, S)$ je gramatika, $A \in N$. Gramatika G obsahuje cyklus, jestliže $A \Rightarrow^+ A$.

Věta 6.2 Jestliže gramatika $G = (N, \Sigma, P, S)$ obsahuje cyklus v nonterminálu A, $A \in N$ a jestliže existuje derivace

$$S \Rightarrow^* \alpha A \beta \Rightarrow^+ w, w \in \Sigma^*, \alpha, \beta \in (N \cup \Sigma)^*$$

pak G je víceznačná.

Důkaz. Existuje-li derivace

$$S \Rightarrow^* \alpha A \beta \Rightarrow^+ w$$

pak vzhledem k existenci cyklu $A \Rightarrow^+$ existuje i derivace

$$S \Rightarrow^* \alpha A \beta \Rightarrow \alpha \gamma_1 \beta \Rightarrow \alpha \gamma_2 \beta \Rightarrow \ldots \Rightarrow \alpha A \beta \Rightarrow^+ w$$

Těmto derivacím přísluší různé derivační stromy. □

Zdroje cyklu

- Jednoduchá pravidla (tvaru $A \rightarrow B$), např.

$$A \Rightarrow B \Rightarrow C \Rightarrow A$$

  v důsledku pravidel $A \rightarrow B$, $B \rightarrow C$, $C \rightarrow A$

- ε-pravidla, např.

$$A \Rightarrow AB \Rightarrow A$$

  v důsledku pravidla $B \rightarrow \varepsilon$

Vlastní gramatika

Definice 6.6 Gramatika bez zbytečných symbolů, ε-pravidel a bez cyklů se nazývá vlastní gramatikou.

Věta 6.3 Každá bezkontextová gramatika má ekvivalentní vlastní gramatiku.

Důkaz. Aplikací algoritmů 4.3 a 4.4 odstraníme zbytečné symboly a ε-pravidla. Jestliže po této transformaci existuje v G derivace $A \Rightarrow^* A$, tj. cyklus, pak jeho příčinou mohou být pouze jednoduchá pravidla a ty lze odstranit aplikací algoritmu 4.5. $\square$

Odstranění levé rekurze

❖ Základem algoritmu je odstranění přímé levé rekurze, tj. levě rekurzivních pravidel, podle následující transformace:

Věta 6.4 Nechť gramatika G má levě rekurzivní pravidla v nonterminálu A a nechť

$$A \to A\alpha_1 \mid A\alpha_2 \mid \ldots \mid A\alpha_m \mid \beta_1 \mid \beta_2 \mid \ldots \mid \beta_n$$

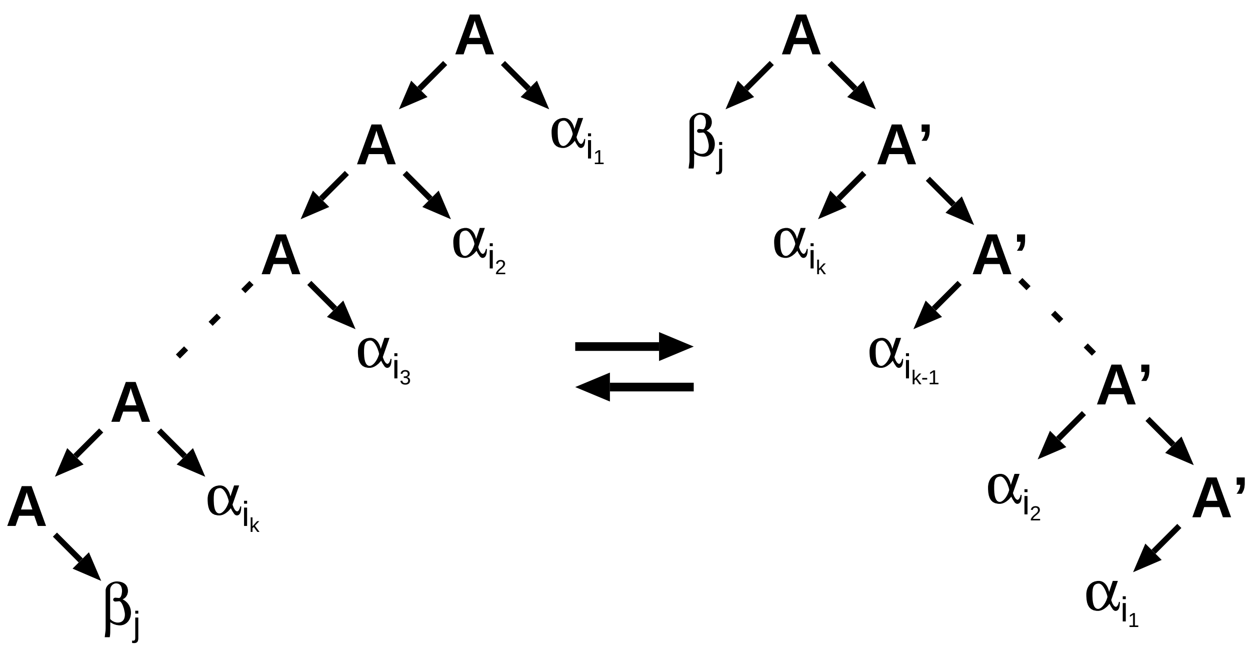
jsou všechna její A-pravidla, přičemž řetězce β_i nezačínají symbolem A. Pak gramatika G', ve které budou tato pravidla nahrazena pravidly:

$$A \to \beta_1 \mid \beta_2 \mid \ldots \mid \beta_n \mid \beta_1 A' \mid \beta_2 A' \mid \ldots \mid \beta_n A'$$

$$A' \to \alpha_1 \mid \alpha_2 \mid \ldots \mid \alpha_1 A' \mid \alpha_2 A' \mid \ldots \mid \alpha_m A'$$

kde A' je nový nonterminál, je ekvivalentní s G.

Důkaz. Uvedená transformace nahrazuje pravidla rekurzivní zleva pravidly, které jsou rekurzivní zprava. Označíme-li jazyky $L_1 = \{\beta_1, \beta_2, \ldots, \beta_n\}$ a $L_2 = \{\alpha_1, \alpha_2, \ldots, \alpha_m\}$, vidíme, že v G lze z nonterminálu A derivovat řetězce tvořící jazyk $L_1 L_2^*$. Právě tyto řetězce můžeme však derivovat z A také v gramatice G'. Efekt popisované transformace ilustruje následující obrázek. $\square$



Příklad 6.6 Uvažujme gramatiku $G = (\{E, T, F\}, \{i, +, *, (,)\}, P, E)$ s pravidly:

$E \rightarrow E + T \mid T$
$T \rightarrow T * F \mid F$
$F \rightarrow (\ E\ ) \mid i$

$$\begin{array}{rcl}
E \rightarrow E + T \mid T & \longrightarrow & E \rightarrow T \mid TE' \\
\alpha_1 = +T, \beta_1 = T & & E' \rightarrow +T \mid\ +TE' \\
T \rightarrow T * F \mid F & \longrightarrow & T \rightarrow F \mid FT' \\
\alpha_1 = *F, \beta_1 = F & & T' \rightarrow *F \mid\ *FT' \\
 & & F \rightarrow (\ E\ ) \mid i
\end{array}$$

Odstranění nepřímé levé rekurze

❖ Odstranění nepřímé levé rekurze spočívá v opakovaném aplikování transformace podle věty 4.1 a transformačních vzorců pro odstranění přímé levé rekurze (věta 4.4).

Příklad 6.7 Uvažujme gramatiku $G = (\{S, A, B\}, \{a, b\}, P, S)$ s pravidly:

$S \to AB$
$A \to BS \mid b$
$B \to SA \mid a$

Na pravidlo $B \to SA$ aplikujeme dvakrát větu 4.1:

$B \to ABA$
$B \to BSBA \mid bBA$

Na všechna B-pravidla

$B \to BSBA \mid bBA \mid a$

aplikujme transformaci věty 4.4:

$B \to bBA \mid a \mid bBAB' \mid aB'$
$B' \to SBA \mid SBAB'$

Normální formy bezkontextových gramatik

Chomského normální forma (CNF)

Definice 6.7 Bezkontextová gramatika $G = (N, \Sigma, P, S)$ je v Chomského normální formě, má-li každé pravidlo z P jeden z těchto tvarů:

1. $A \to BC$, kde $A, B, C \in N$
2. $A \to a$, kde $a \in \Sigma$
3. je-li $\varepsilon \in L(G)$, pak $S \to \varepsilon$ je jediné ε-pravidlo a S se nevyskytuje na pravé straně žádného přepisovacího pravidla.

Problém: Nechť $G = (N, \Sigma, P, S)$ je bezkontextová gramatika v CNF a nechť $w \in L(G)$ a $S \Rightarrow^p_G w$. Jaká je délka řetězce w?

Řešení: Označme $|w| = n$. Zřejmě platí

$$p = n + (n-1) = 2n - 1$$

$$|w| = \frac{p+1}{2}$$

Věta 6.5 Nechť G je bezkontextová gramatika. Pak existuje gramatika G' v Chomského normální formě taková, že $L(G') = L(G)$.

Důkaz. (Hlavní myšlenka) Gramatiku G převedeme na ekvivalentní vlastní gramatiku bez jednoduchých pravidel.

1. Pravidla tvaru (1), (2) a (3) ponecháme.
2. Pravidla tvaru $A \rightarrow X_1X_2 \dots X_n$, kde $X_i \in (N \cup \Sigma)$ pro $i = 1, \dots, n$, $n > 2$, transformujeme na $A \rightarrow X_1'\langle X_2X_3 \dots X_n\rangle$, kde $\langle X_2X_3 \dots X_n\rangle$ je nový nonterminál a X_1' je nový nonterminál pokud $X_1 \in \Sigma$, nebo $X_1' = X_1$ v opačném případě.
3. Pravidla tvaru $A \rightarrow X_1X_2$ transformujeme na pravidla $A \rightarrow X_1'X_2'$, kde X_i' je nový nonterminál pokud $X_i \in \Sigma$, nebo $X_i' = X_i$ v opačném případě pro $i \in \{1, 2\}$
4. Pro nové nonterminály tvaru $\langle X_1X_2 \dots X_n\rangle$, $n \geq 2$, zavedeme pravidla $\langle X_1X_2 \dots X_n\rangle \rightarrow X_1'\langle X_2 \dots X_n\rangle$ pro $n > 2$ a $\langle X_1X_2\rangle \rightarrow X_1'X_2'$ pro $n = 2$, kde $\langle X_2 \dots X_n\rangle$ je nový nonterminál a X_i' je nový nonterminál pokud $X_i \in \Sigma$, nebo $X_i' = X_i$ v opačném případě pro $i \in \{1, 2\}$.
5. Pro nové nonterminály tvaru X_i', kde $X_i \in \Sigma$ přidáme pravidla tvaru $X_i' \rightarrow X_i$.

$\square$

Příklad 6.8 Uvažujme gramatiku $G = (\{A, B\}, \{a, b, c\}, P, A)$ s pravidly:

$A \rightarrow BAB \mid Ba \mid bc$
$B \rightarrow AB \mid a \mid BBB$

Po aplikaci transformací (1.)-(4.) získáme CNF ve tvaru:

$A \rightarrow B\langle AB\rangle|Ba'|b'c'$
$B \rightarrow AB \mid a \mid B\langle BB\rangle$
$\langle AB\rangle \rightarrow AB$
$\langle BB\rangle \rightarrow BB$
$a' \rightarrow a$
$b' \rightarrow b$
$c' \rightarrow c$

Greibachové normální forma (GNF)

Definice 6.8 Bezkontextová gramatika $G = (N, \Sigma, P, S)$ je v Greibachové normální formě, je-li G gramatikou bez ε-pravidel a každé pravidlo z P (vyjma případného pravidla $S \to \varepsilon$) má tvar:
$A \to a\alpha$, kde $a \in \Sigma, \alpha \in N^*$

Lemma 6.1 Nechť $G = (N, \Sigma, P, S)$ je bezkontextová gramatika bez levé rekurze. Pak na N existuje lineární uspořádání $\prec$ takové, že
je-li $A \to B\alpha$ pravidlo z P, pak $A \prec B$.

Důkaz. Definujme relaci R na N:
$R = \{(A, B) | A \Rightarrow^* B\alpha, A, B \in N, \alpha \in (N \cup \Sigma)^*\}$
Lze ukázat, že R je částečné uspořádání. Každé částečné uspořádání lze rozšířit na lineární uspořádání. $\square$

Věta 6.6 Ke každé bezkontextové gramatice existuje Greibachové normální forma této gramatiky.

Důkaz. (Hlavní myšlenka) Nalezení GNF gramatiky G předpokládá:

1. Odstranění levé rekurze a ε-pravidel.
2. Nalezení lineárního uspořádání $\prec$ na množině nonterminálů.
3. Aplikace substitucí v pořadí opačném k danému uspořádání $\prec$ tak, aby všechna pravidla byla tvaru $A \to a\beta, A \in N, a \in \Sigma, \beta \in (N \cup \Sigma)^*$.
4. Převod všech pravidel $A \to a\beta, A \in N, \beta \in (N \cup \Sigma)^*$ na pravidla tvaru $A \to a\alpha$, $A \in N, a \in \Sigma$, $\alpha \in N^*$.

$\square$

Příklad 6.9 Uvažujme gramatiku, jež vznikne odstraněním levé rekurze z typické gramatiky pro aritmetický výraz $G = (\{E, T, F\}, \{i, +, *, (,)\}, P, E)$ s pravidly:

$E \to T \mid TE'$
$E' \to +T \mid +TE'$
$T \to F \mid FT'$
$T' \to *F \mid *FT'$
$F \to (\ E\ ) \mid i$

Nalezneme lineární uspořádání: $E' \prec E \prec T' \prec T \prec F$

$E \to (\ E\ )' \mid i \mid (\ E\ )'T' \mid iT' \mid (\ E\ )'E' \mid iE' \mid (\ E\ )'T'E' \mid iT'E'$
$E' \to +T \mid +TE'$
$T \to (\ E\ )' \mid i \mid (\ E\ )'T' \mid iT'$
$T' \to *F \mid *FT'$
$F \to i \mid (\ E\ )'$

Přidáno pravidlo $)' \to)$.