Kyushu University 100th 2011
知の新世紀を拓く

形式言語理論
Formal Language Theory

9.　文脈自由文法

九州大学
KYUSHU UNIVERSITY

本日の内容

- 準備
- 文脈自由文法(CFG)
- 導出木・最左導出
- 正規文法
- 線形文法

準備（復習）

復習

- A を集合とする. $A \times A$ の部分集合を A 上の**関係**と呼ぶ.
- A 上の関係 R, S に対して, R, S の**積** $R{\cdot}S$ を次で定義する.
  - $R \cdot S = \{\, (a,c) \mid (a,b)\in R \wedge (b,c)\in S \,\}$
- $R{\cdot}S$ もまた A 上の関係である.
  - $R \cdot S$ を単に RS と書くことが多い.

復習

- $I_A = \{(a,a) \mid a \in A\}$ とおく.
- A 上の任意の関係 R に対し次が成り立つ.

$$R\,I_A = I_A R = R$$

- A 上の関係 R に対して, R^n $(n = 0, 1,2,\ldots)$を次で定める.
  - $R^0 = I_A, \quad R^n = RR^{n-1} \ (n \geq 1)$
- R の**推移閉包** R^+ と**反射的推移閉包** R^*を次で定める.
  - $R^+ = \bigcup_{n=1}^{\infty} R^n$
  - $R^* = \bigcup_{n=0}^{\infty} R^n$

文脈自由文法

文脈自由文法(context-free grammar; CFG)

【定義】文脈自由文法とは，次のような4つ組
$G = (N, \Sigma, P, S)$ をいう.

- N は非終端記号の空でない有限集合.
- Σ は終端記号の空でない有限集合.
- $P \subseteq N \times (N \cup \Sigma)^*$ は生成規則の集合.
- $S \in N$ は開始記号.

各生成規則 $(A, \alpha) \in N \times (N \cup \Sigma)^*$ を以下のように書く.

$$A \rightarrow \alpha$$

例

$$G_1 = (N, \Sigma, P, S)$$
$$N = \{S\}$$
$$\Sigma = \{a,\ b\}$$
$$P = \{S \rightarrow ab,\ S \rightarrow aSb\}$$

複数の生成規則をまとめて書く

- 同じ非終端記号Aを左辺にもつ生成規則

  $$A \rightarrow \alpha_1, \ldots, A \rightarrow \alpha_n$$

  をまとめて次のように書くことがある.

  $$A \rightarrow \alpha_1 \mid \cdots \mid \alpha_n$$

例

$$G_2 = (N, \Sigma, P, S)$$

$$N = \{S\}$$

$$\Sigma = \{(,\ )\}$$

$$P = \{S \rightarrow SS,\ S \rightarrow (S),\ S \rightarrow ()\}$$

$$P = \{S \rightarrow SS \mid (S) \mid ()\}$$

と書いてもよい

例

$$G_3 = (N, \Sigma, P, S)$$

$$N = \{S, A\}$$

$$\Sigma = \{0, 1\}$$

$$P = \{S \rightarrow 0,\ S \rightarrow 1A,\ A \rightarrow \varepsilon,\ A \rightarrow 0A,\ A \rightarrow 1A\}$$

$$P = \{S \rightarrow 0 | 1A,\ A \rightarrow \varepsilon | 0A | 1A\}$$

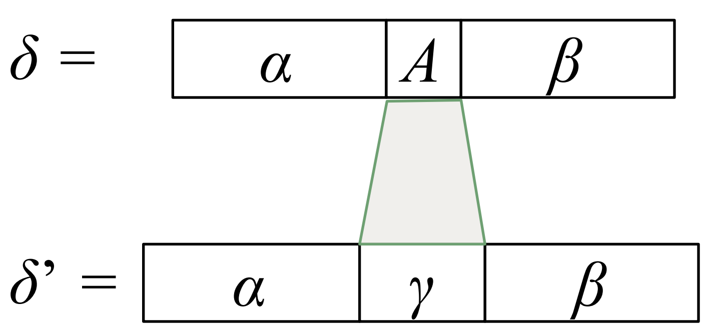
導出

【定義】 CFG $G = (N, \Sigma, P, S)$ に対し，$(N \cup \Sigma)^*$ 上の2項関係 $\underset{G}{\Rightarrow}$ を次で定義する.

$$\delta \underset{G}{\Rightarrow} \delta'$$

$$\Leftrightarrow$$

$$\exists \alpha,\ \exists \beta \in (N \cup \Sigma)^*,\ \exists A \to \gamma \in P,$$
$$\delta = \alpha A \beta \wedge \delta' = \alpha \gamma \beta$$



例 導出

$$G_1 = (N, \Sigma, P, S)$$
$$N = \{S\}$$
$$\Sigma = \{a, b\}$$
$$P = \{S \to ab \mid aSb\}$$

$$S \underset{G_1}{\Rightarrow} ab$$

$$S \underset{G_1}{\Rightarrow} aSb \underset{G_1}{\Rightarrow} aabb$$

$$S \underset{G_1}{\Rightarrow} aSb \underset{G_1}{\Rightarrow} aaSbb \underset{G_1}{\Rightarrow} aaabbb$$

$$\vdots$$

例 導出

$$G_2 = (N, \Sigma, P, S)$$
$$N = \{S\}$$
$$\Sigma = \{(,)\}$$
$$P = \{S \to SS,\ S \to (S),\ S \to ()\}$$

$$S \underset{G_2}{\Rightarrow} (S) \underset{G_2}{\Rightarrow} (())$$

$$S \underset{G_2}{\Rightarrow} SS \underset{G_2}{\Rightarrow} SSS \underset{G_2}{\Rightarrow} ()SS \underset{G_2}{\Rightarrow} ()(S)S \underset{G_2}{\Rightarrow} ()(())S \underset{G_2}{\Rightarrow} ()(())()$$

$$S \underset{G_2}{\Rightarrow} SS \underset{G_2}{\Rightarrow} (S)S \underset{G_2}{\Rightarrow} (S)() \underset{G_2}{\Rightarrow} (())()$$

$$\vdots$$

例 導出

$$G_3 = (N, \Sigma, P, S)$$
$$N = \{S, A\}$$
$$\Sigma = \{0, 1\}$$
$$P = \{S \rightarrow 0,\ S \rightarrow 1A,\ A \rightarrow \varepsilon,\ A \rightarrow 0A,\ A \rightarrow 1A\}$$

$$S \underset{G_3}{\Rightarrow} 0$$

$$S \underset{G_3}{\Rightarrow} 1A \underset{G_3}{\Rightarrow} 1$$

$$S \underset{G_3}{\Rightarrow} 1A \underset{G_3}{\Rightarrow} 10A \underset{G_3}{\Rightarrow} 101A \underset{G_3}{\Rightarrow} 1011A \underset{G_3}{\Rightarrow} 1011$$

$$\vdots$$

導出

- $G = (N, \Sigma, P, S)$ を CFG とする.
- $\alpha, \beta \in (N \cup \Sigma)^*$ に対して:
  - $\alpha \underset{G}{\Rightarrow}^{n} \beta$ であるとき,
    α から β が n ステップで導出されるという.
  - $\alpha \underset{G}{\Rightarrow}^{*} \beta$ であるとき,
    α から β が導出されるという.

生成する言語

【定義】 CFG $G=(N, \Sigma, P, S)$ と文字列 $w \in \Sigma^*$ に対して，G が w を**生成する**とは，$S \underset{G}{\Rightarrow}^* w$ であるときをいう．

【定義】 CFG $G=(N, \Sigma, P, S)$ によって生成される Σ 上の文字列全体の集合を $L(G)$ で表わし，**G の生成する言語**とよぶ．すなわち，

$$L(G)=\{\, w \mid w \in \Sigma^* \wedge S \underset{G}{\Rightarrow}^* w \}$$

である．

文脈自由言語（context-free language; CFL）

【定義】 CFG G と言語 $L \subseteq \Sigma^*$ に対し
$L = L(G)$ となるとき，G は L を生成するという.

【定義】 言語 $L \subseteq \Sigma^*$ が文脈自由言語であるとは，
L を生成する CFG G が存在するときをいう.

例 文脈自由言語

$$
\boxed{
\begin{aligned}
&G_1 = (N, \Sigma, P, S)\\
&\quad N = \{S\}\\
&\quad \Sigma = \{a,\ b\}\\
&\quad P = \{S \to ab,\ S \to aSb\}
\end{aligned}
}
$$

$$S \underset{G_1}{\Rightarrow} ab$$

$$S \underset{G_1}{\Rightarrow} aSb \underset{G_1}{\Rightarrow} aabb$$

$$S \underset{G_1}{\Rightarrow} aSb \underset{G_1}{\Rightarrow} aaSbb \underset{G_1}{\Rightarrow} aaabbb$$

$$\vdots$$

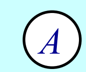
$$\boxed{L(G_1) = \{a^n b^n \mid n \geq 1\}}$$

導出木・最左導出

導出木 (または構文木)

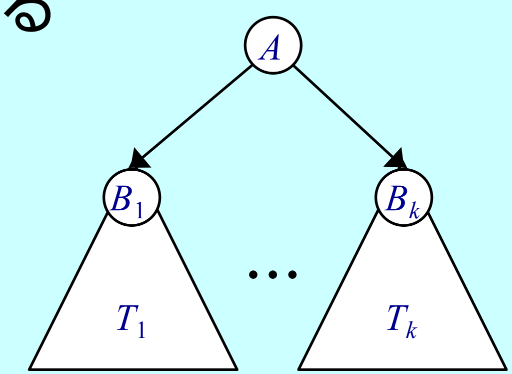
【定義】 CFG $G = (N, \Sigma, P, S)$ に対する導出木とは，次で定まる頂点ラベル付き木をいう.

(1) 各 $A \in N \cup \Sigma$ に対して，記号 A をラベルとする頂点1個から成る木は導出木である.



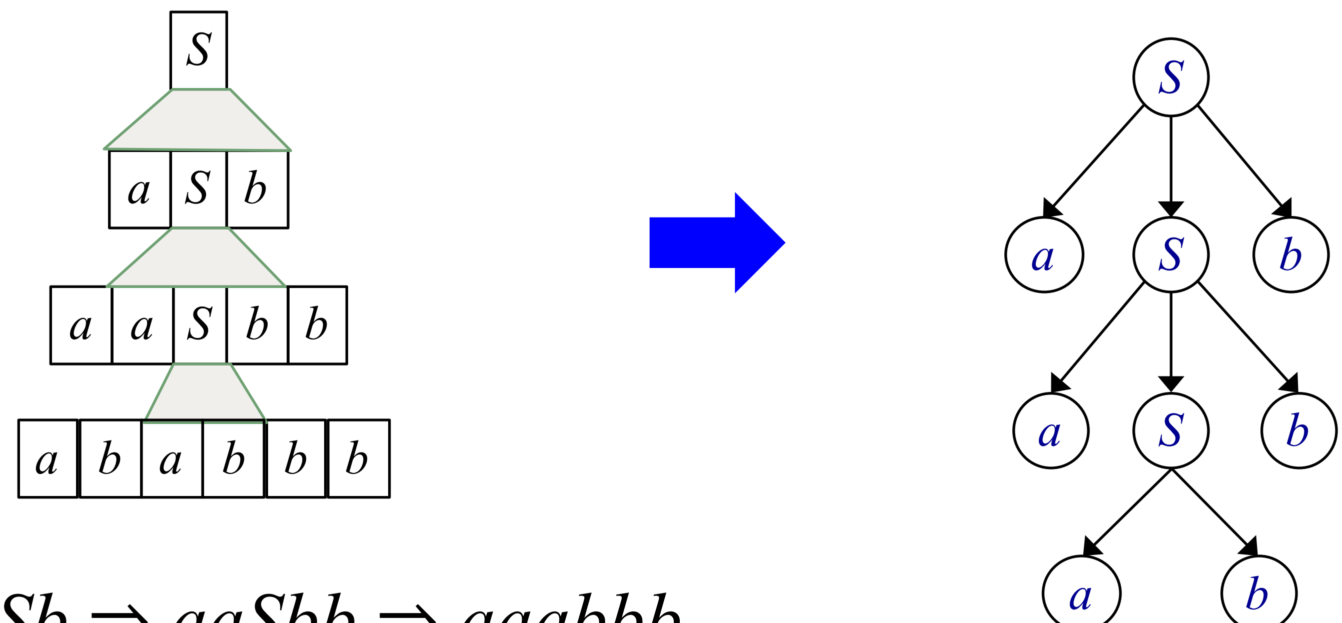
(2) 各生成規則 $A \rightarrow \varepsilon \in P$ に対して，記号 ε をラベルとする頂点を子とし記号 A をラベルとする頂点を根とする木は導出木である.

(3) $T_1, \ldots, T_k$ を根のラベルが $B_1, \ldots, B_k \in N \cup \Sigma$ である導出木とする.
生成規則 $A \rightarrow B_1 \ldots B_k \in P$ に対して，
$T_1, \ldots, T_k$ の根を子としラベル A をもつ頂点を根とする木は導出木である.



例 導出木

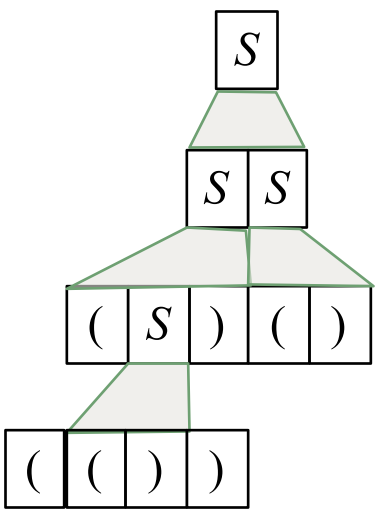
$$G_1 = (N, \Sigma, P, S)$$
$$N = \{S\}$$
$$\Sigma = \{a, b\}$$
$$P = \{S \rightarrow ab,\ S \rightarrow aSb\}$$

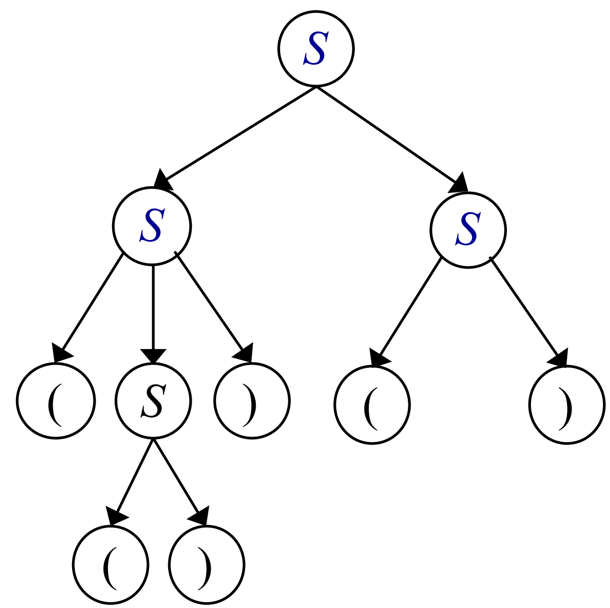


$$S \underset{G_1}{\Rightarrow} aSb \underset{G_1}{\Rightarrow} aaSbb \underset{G_1}{\Rightarrow} aaabbb$$

例 導出木

$$G_2 = (N, \Sigma, P, S)$$
$$N = \{S\}$$
$$\Sigma = \{(,\ )\}$$
$$P = \{S \rightarrow SS,\ S \rightarrow (S),\ S \rightarrow ()\}$$



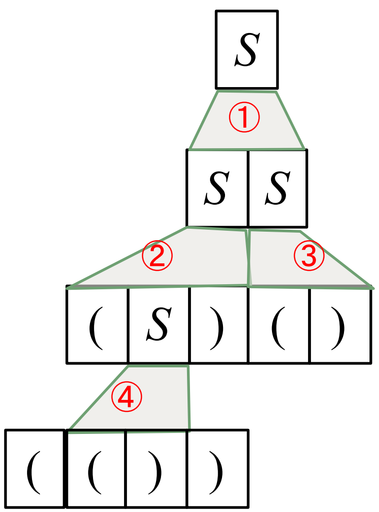


$$S \underset{G_2}{\Rightarrow} SS \underset{G_2}{\Rightarrow} (S)S \underset{G_2}{\Rightarrow} (S)() \underset{G_2}{\Rightarrow} (())()$$

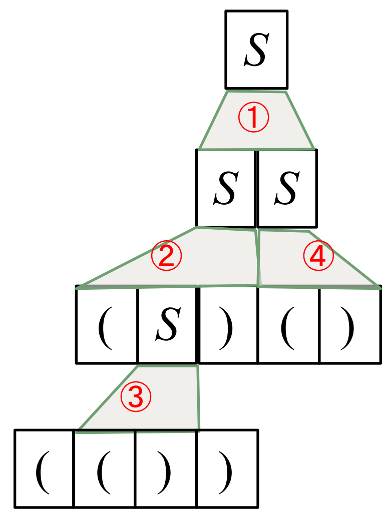
最左導出

【定義】 $G=(N,\Sigma,P,S)$ を CFG とし，$\alpha,\beta\in(N\cup\Sigma)^*$ とする．導出 $\alpha \underset{G}{\Rightarrow}^* \beta$ が**最左導出**であるとはその導出の各ステップにおいて，最も左にある非終端記号が置換されているときをいう．

× $S \underset{G_2}{\Rightarrow} SS \underset{G_2}{\Rightarrow} (S)S \underset{G_2}{\Rightarrow} (S)() \underset{G_2}{\Rightarrow} (())()$



○ $S \underset{G_2}{\Rightarrow} SS \underset{G_2}{\Rightarrow} (S)S \underset{G_2}{\Rightarrow} (())S \underset{G_2}{\Rightarrow} (())()$



最左導出に関する命題

【命題】 $G=(N,\Sigma,P,S)$を CFG とし, $\alpha,\beta\in(N\cup\Sigma)^*$ とする.
導出 $\alpha \underset{G}{\Rightarrow}^* \beta$ が存在するならば
α から β への最左導出が存在する.

正規文法

正規文法

【定義】正規文法とは，生成規則が次のいずれかの形式に制限された CFG $G = (N, \Sigma, P, S)$ をいう.

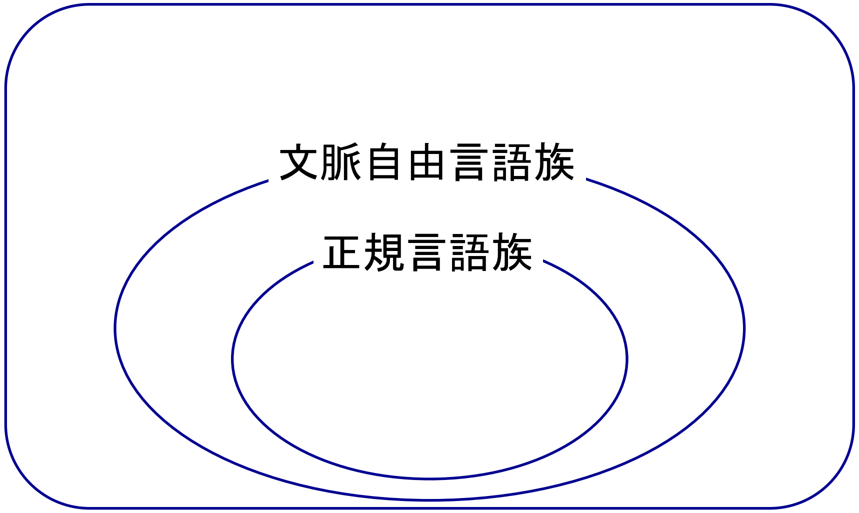
1. $A \rightarrow aB \quad (A, B \in N,\ a \in \Sigma)$
2. $A \rightarrow a \quad (A \in N,\ a \in \Sigma)$
3. $S \rightarrow \varepsilon$

正規文法と正規言語

【定理4.1】 任意の言語 $L \subseteq \Sigma^*$ に対し次の(1)(2)は等価である.

(1) L は正規言語である.

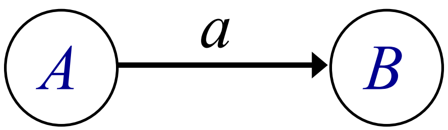
(2) 正規文法 G が存在して $L=L(G)$ である.



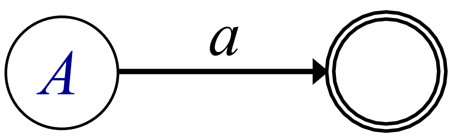
定理4.1の証明：概略

- 正規文法の生成規則とε-動作付きNFAの状態遷移を次のように対応させればよい.

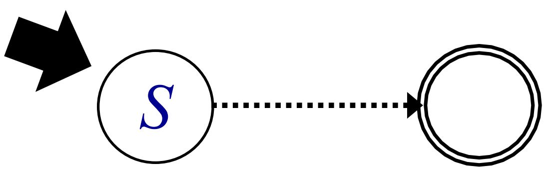
$A \to aB$ ⟺



$A \to a$ ⟺

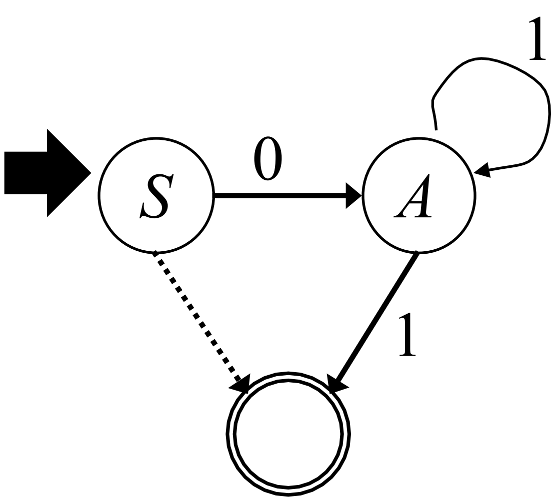


$S \to \varepsilon$ ⟺



例

$$G_4 = (N, \Sigma, P, S)$$
$$N = \{S, A\}$$
$$\Sigma = \{0, 1\}$$
$$P = \{S \to \varepsilon, S \to 0A, A \to 1, A \to 1\mathrm{A}\}$$



線形文法

右線形文法・左線形文法

【定義】右線形文法とは，生成規則が次のいずれかの形式に制限された CFG $G = (N, \Sigma, P, S)$ をいう．

1. $A \rightarrow uB \quad (A, B \in N,\ u \in \Sigma^*)$.
2. $A \rightarrow u \quad (A \in N,\ u \in \Sigma^*)$.

【定義】左線形文法とは，生成規則が次のいずれかの形式に制限された CFG $G = (N, \Sigma, P, S)$ をいう．

1. $A \rightarrow Bu \quad (A, B \in N,\ u \in \Sigma^*)$.
2. $A \rightarrow u \quad (A \in N,\ u \in \Sigma^*)$.

線形文法と正規言語

【定理4.2】 任意の言語 $L \subseteq \Sigma^*$ に対し次の(1)(2)(3)は等価である.

(1) L は正規言語である.

(2) 右線形文法 G が存在して $L=L(G)$ である.

(3) 左線形文法 G が存在して $L=L(G)$ である.

右線形文法と正規文法

【補題4.1】任意の右線形文法 G に対し，$L(G)=L(G')$を満たし，生成規則が次のいずれかの形式に制限された右線形文法 G' が存在する.

① $A \rightarrow uB\ (A,B \in N,\ u \in \Sigma^{+})$.

② $A \rightarrow u\ \ (A \in N,\ u \in \Sigma^{+})$.

③ $S \rightarrow \varepsilon$

補題4.1の証明

- 右線形文法 $G = (N, \Sigma, P, S)$ に対して，次のような右線形文法 $G' = (N, \Sigma, P', S)$ は補題の条件を満たす．
  - $P' = \{A \to uC \mid u \in \Sigma^+, A \underset{G}{\Rightarrow}^* B, B \to uC \in P\}$
    $\cup \{A \to u \mid u \in \Sigma^+, A \underset{G}{\Rightarrow}^* B, B \to uC \in P, C \underset{G}{\Rightarrow}^* \varepsilon\}$
    $\cup \{A \to u \mid u \in \Sigma^+, A \underset{G}{\Rightarrow}^* B, B \to u \in P\}$
    $\cup \{S \to \varepsilon \mid S \underset{G}{\Rightarrow}^* \varepsilon\}$
- $L(G)=L(G')$が成り立つ．

QED

定理4.2の証明: (1)⇔(2)

- 定理4.1と定義より(1)⇒(2)は明らか. (2)⇒(1)を示す.
- Gを任意の右線形文法とする. 補題4.1より, Gの生成規則は①②③のいずれかの形式としてよい.
- Gの生成規則を以下のように置換すればよい.

$$\boxed{\pi = A \to a_1 \dots a_{m+1} B \ (m \geq 1)}$$

$$\Downarrow$$

$$\boxed{\begin{array}{l} A \to a_1 A_{\pi,1}, \\ A_{\pi,1} \to a_2 A_{\pi,2}, \\ \dots \\ A_{\pi,m-1} \to a_m A_{\pi,m} \\ A_{\pi,m} \to a_{m+1} B \end{array}}$$

$$\boxed{\pi = A \to a_1 \dots a_{m+1} \ (m \geq 1)}$$

$$\Downarrow$$

$$\boxed{\begin{array}{l} A \to a_1 A_{\pi,1}, \\ A_{\pi,1} \to a_2 A_{\pi,2}, \\ \dots \\ A_{\pi,m-1} \to a_m A_{\pi,m} \\ A_{\pi,m} \to a_{m+1} \end{array}}$$

定理4.2の証明: (1)⇔(3)

- 任意の言語 $L \subseteq \Sigma^*$ に対し，$L^R=\{w^R | w \in L\}$とおく．任意の CFG G に対し，生成規則の右辺を反転させて得られる CFG を G^R で表す．次が成り立つ．
  - L が正規 ⇔ L^R が正規.
  - $L(G^R)=L(G)^R$.
  - 任意の左線形文法 G に対して G^R は右線形文法.
- 次の関係から(1)⇔(3)が成り立つ．
  - Lが正規
⇔L^Rが正規
⇔右線形文法 G が存在して$L^R=L(G)$
⇔左線形文法 G'が存在して$L=L(G')$.