Kyushu University 100th 2011
知の新世紀を拓く

形式言語理論

Formal Language Theory

13.　句構造文法とChomsky階層

九州大学
KYUSHU UNIVERSITY

本日の内容

- Chomsky階層
- 句構造文法
- 文脈依存文法
- Turing機械
- 線形有界オートマトン
- 各言語族の閉性

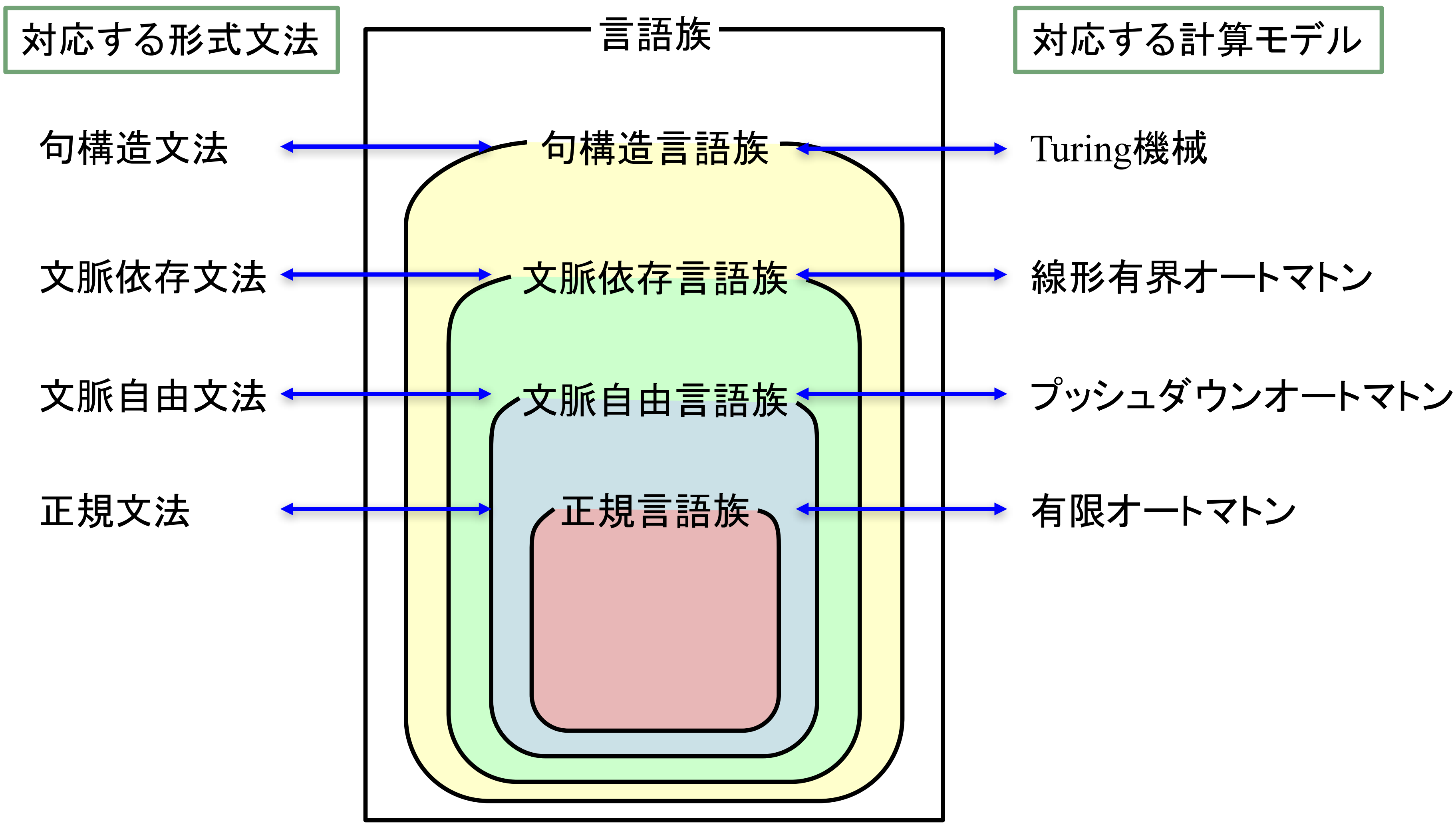
Chomsky階層

Avram Noam Chomsky (1928-)

- 合衆国の哲学者，言語哲学者，言語学者，認知科学者，論理学者．
- 「現代言語学の父」

Wikipediaより引用

Chomsky階層


対応する形式文法
言語族
対応する計算モデル
句構造文法
句構造言語族
Turing機械
文脈依存文法
文脈依存言語族
線形有界オートマトン
文脈自由文法
文脈自由言語族
プッシュダウンオートマトン
正規文法
正規言語族
有限オートマトン

句構造文法

句構造文法(phrase-structure grammar; PSG)

【定義】句構造文法とは，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.
- $S \in N$ は開始記号.
- P は以下の形式をした生成規則の有限集合

$$\alpha \to \beta \quad (\alpha \in (N \cup \Sigma)^* N (N \cup \Sigma)^*, \beta \in (N \cup \Sigma)^*)$$

左辺 α は少なくとも1個の非終端記号を含む

文脈自由文法は句構造文法の特別な場合である

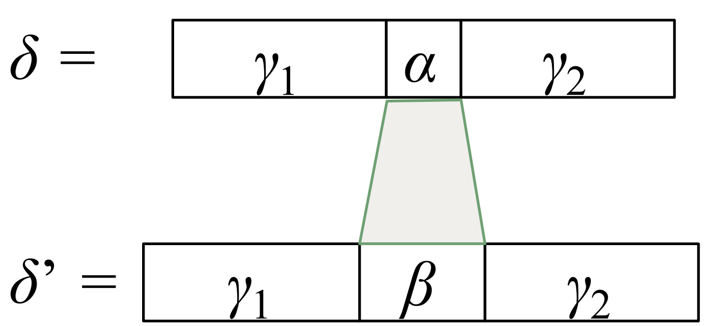
句構造文法における導出

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

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

$$\Leftrightarrow$$

$$\exists\gamma_1,\ \exists\gamma_2\in(N\cup\Sigma)^*,\ \exists\alpha\rightarrow\beta\in P,$$
$$\delta=\gamma_1\,\alpha\,\gamma_2 \wedge \delta'=\gamma_1\,\beta\,\gamma_2$$



例 句構造文法

$$L(G_1)=\{a^{2^n} \mid n\geq 0\}=\{a^1, a^2, a^4, a^8, \ldots\}$$

$$
\begin{aligned}
G_1 &= (N, \Sigma, P, S)\\
&N=\{S, A, B, C\}\\
&\Sigma=\{a\}\\
&P=\{\\
&\qquad S\rightarrow BAB,\\
&\qquad BA \rightarrow BC,\\
&\qquad CA \rightarrow AAC,\\
&\qquad CB \rightarrow AAB,\\
&\qquad A \rightarrow a,\\
&\qquad B \rightarrow \varepsilon\\
&\}
\end{aligned}
$$

$$S \Rightarrow BAB$$

$$\underline{BA}B \Rightarrow B\underline{CB} \Rightarrow BAAB$$

$$\vdots$$

$$\Rightarrow^* BAAB$$

$$\underline{BA}AB \Rightarrow B\underline{CA}B \Rightarrow BAA\underline{CB} \Rightarrow BAAAAB$$

$$\vdots$$

$$\Rightarrow^* BAAAAB$$

$$\vdots$$

$$\Rightarrow^* BAAAAAAAAB$$

$$\Rightarrow^* aaaaaaaa$$

例 句構造文法

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

$$N = \{S, A, B, C, D, E\}$$

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

$$P = \{$$

$$S \rightarrow ABC,$$

$$AB \rightarrow 0AD | 1AE,$$

$$DC \rightarrow B0C, \quad EC \rightarrow B1C,$$

$$D0 \rightarrow 0D, \quad D1 \rightarrow 1D,$$

$$E0 \rightarrow 0E, \quad E1 \rightarrow 1E,$$

$$AB \rightarrow \varepsilon, \quad C \rightarrow \varepsilon,$$

$$0B \rightarrow B0, \quad 1B \rightarrow B1$$

$$\}$$

$$L(G_2) = \{uu \mid u \in \{0,1\}^*\}$$

$$S \Rightarrow^* uABuC \Rightarrow^* uu$$

$u=x0$とする.

$$S \Rightarrow^* xABxC$$
$$\Rightarrow x0ADxC$$
$$\Rightarrow^* x0AxDC$$
$$\Rightarrow x0AxB0C$$
$$\Rightarrow^* x0ABx0C$$
$$= uABuC$$

$u=x1$とする.

$$S \Rightarrow^* xABxC$$
$$\Rightarrow x1AExC$$
$$\Rightarrow^* x1AxEC$$
$$\Rightarrow x1AxB1C$$
$$\Rightarrow^* x1ABx1C$$
$$= uABuC$$

句構造文法の生成する言語

- 文脈自由文法の場合と同様，以下のように定める．

【定義】 PSG $G = (N, \Sigma, P, S)$ 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】 PSG $G = (N, \Sigma, P, S)$ 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，

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

である．

句構造言語

【定義】言語 $L \subseteq \Sigma^*$ が**句構造言語**であるとは，$L = L(G)$ となる句構造文法 G が存在するときをいう.

0型文法，消去付き文脈依存文法

【定義】句構造文法 $G=(N, \Sigma, P, S)$が0型であるとは，すべての生成規則が次の形式であるときをいう．

$$\alpha \to \beta \quad (\alpha \in N^{+}, \beta \in (N \cup \Sigma)^{*})$$

【定義】句構造文法 $G=(N, \Sigma, P, S)$が消去付文脈依存文法であるとは，すべての生成規則が次の形式であるときをいう．

$$\alpha A \beta \to \alpha \gamma \beta \quad (A \in N, \gamma \in (N \cup \Sigma)^{*}, \alpha, \beta \in (N \cup \Sigma)^{*})$$

句構造言語

【定理7.1】 任意の言語 $L \subseteq \Sigma^*$ に対し，次の①②③は等価である.

① $L=L(G)$ となる句構造文法 G が存在する.

② $L=L(G)$ となる0型文法 G が存在する.

③ $L=L(G)$ となる消去付文脈依存文法 G が存在する.

文脈依存文法

文脈依存文法(context-sensitive grammar; CSG)

【定義】句構造文法 $G=(N, \Sigma, P, S)$ が**文脈依存文法**であるとは, P の生成規則が次のいずれかの形式であるときをいう.

(1) $S \longrightarrow \varepsilon$

(2) $\alpha A \beta \longrightarrow \alpha\gamma\beta \quad (A \in N, \gamma \in (N \cup \Sigma)^{+}, \alpha, \beta \in (N \cup \Sigma)^{*})$

- 文脈自由文法における生成規則 $A \rightarrow \gamma$ と異なり, 記号 A が**左右の文脈 α, β** の中にあるときのみ書き換えが行われる.

例 文脈依存文法

$$L(G_3)=\{a^n b^n c^n \mid n > 0\}$$

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

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

$$\Sigma = \{a, b, c\}$$

$$P = \{$$

$$S \to aSBC \mid aBC,$$

$$CB \to AB,$$

$$AB \to AC,$$

$$AC \to BC,$$

$$aB \to ab,\ bB \to bb,$$

$$bC \to bc,\ cC \to cc$$

$$\}$$

$$S \Rightarrow^* a^n(BC)^n$$

$$\vdots$$

$$\Rightarrow^* a^n B^n C^n$$

$$\vdots$$

$$\Rightarrow^* a^n b^n C^n$$

$$\vdots$$

$$\Rightarrow^* a^n b^n c^n$$

$$\begin{aligned} & (BC)^n \\ = \ & B(CB)^{n-1}C \\ \Rightarrow^* \ & B(BC)^{n-1}C \\ \Rightarrow^* \ & B(B^{n-1}C^{n-1})C \\ = \ & B^nC^n \end{aligned}$$

$$\begin{aligned} & aB^n \\ = \ & aB^{n-1}B \\ \Rightarrow^* \ & ab^{n-1}B \\ \Rightarrow^* \ & ab^{n-1}b \\ = \ & ab^n \end{aligned}$$

$$\begin{aligned} & bC^n \\ = \ & bC^{n-1}C \\ \Rightarrow^* \ & bc^{n-1}C \\ \Rightarrow^* \ & bc^{n-1}c \\ = \ & bc^n \end{aligned}$$

文脈依存言語

【定義】 言語 $L \subseteq \Sigma^*$ が文脈依存言語であるとは,
$L=L(G)$ となる文脈依存文法 G が存在するときをいう.

単調文法

【定義】句構造文法 $G = (N, \Sigma, P, S)$ が単調文法であるとは，任意の生成規則 $\alpha \rightarrow \beta$ が次のいずれかを満たすときをいう.

(1) $|\alpha| \leq |\beta|$

(2) $\alpha = S, \beta = \varepsilon$

例 単調文法

$G_4 = (N, \Sigma, P, S)$

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

$\Sigma = \{a, b, c\}$

$P = \{$

$S \to aSBC,\ S \to aBC,$

$CB \to BC,$

$aB \to ab,\ bB \to bb,$

$bC \to \mathrm{bc},\ cC \to cc$

$\}$

$L(G_4) = \{a^n b^n c^n \mid n > 0\}$

G_3 の生成規則
$CB \to AB,\ AB \to AC,\ AC \to BC$
の代わり

文脈依存言語

【定理7.2】 任意の言語 $L \subseteq \Sigma^*$ に対し，次の①②は等価である.

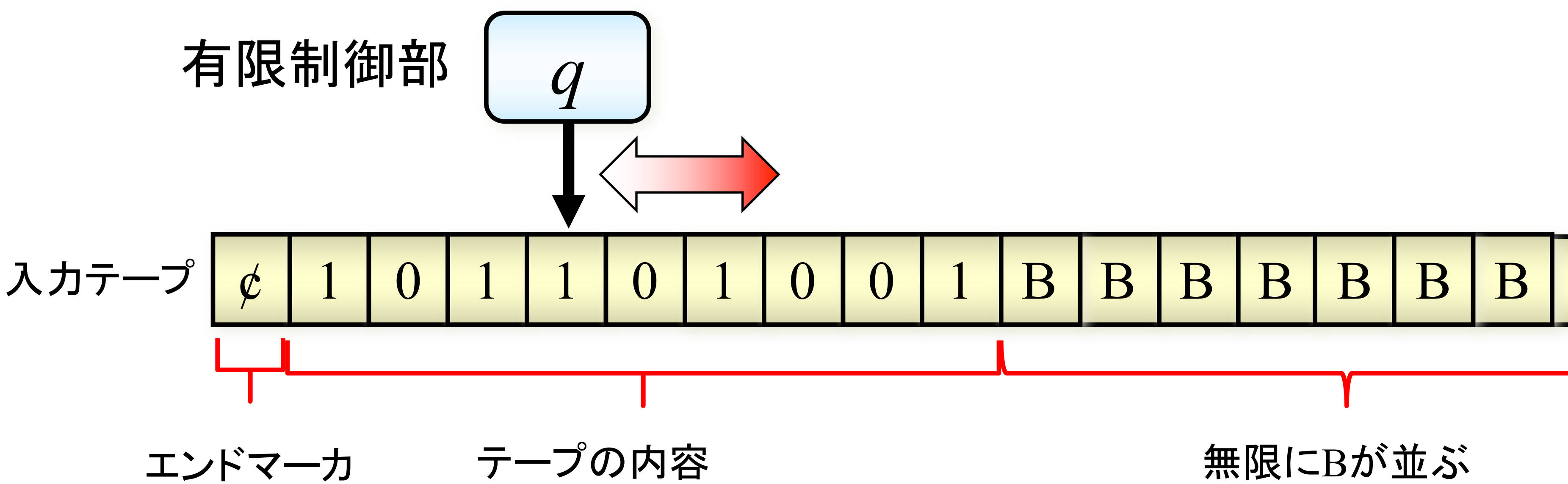
① $L=L(G)$ となる文脈依存文法 G が存在する.

② $L=L(G)$ となる単調文法 G が存在する.

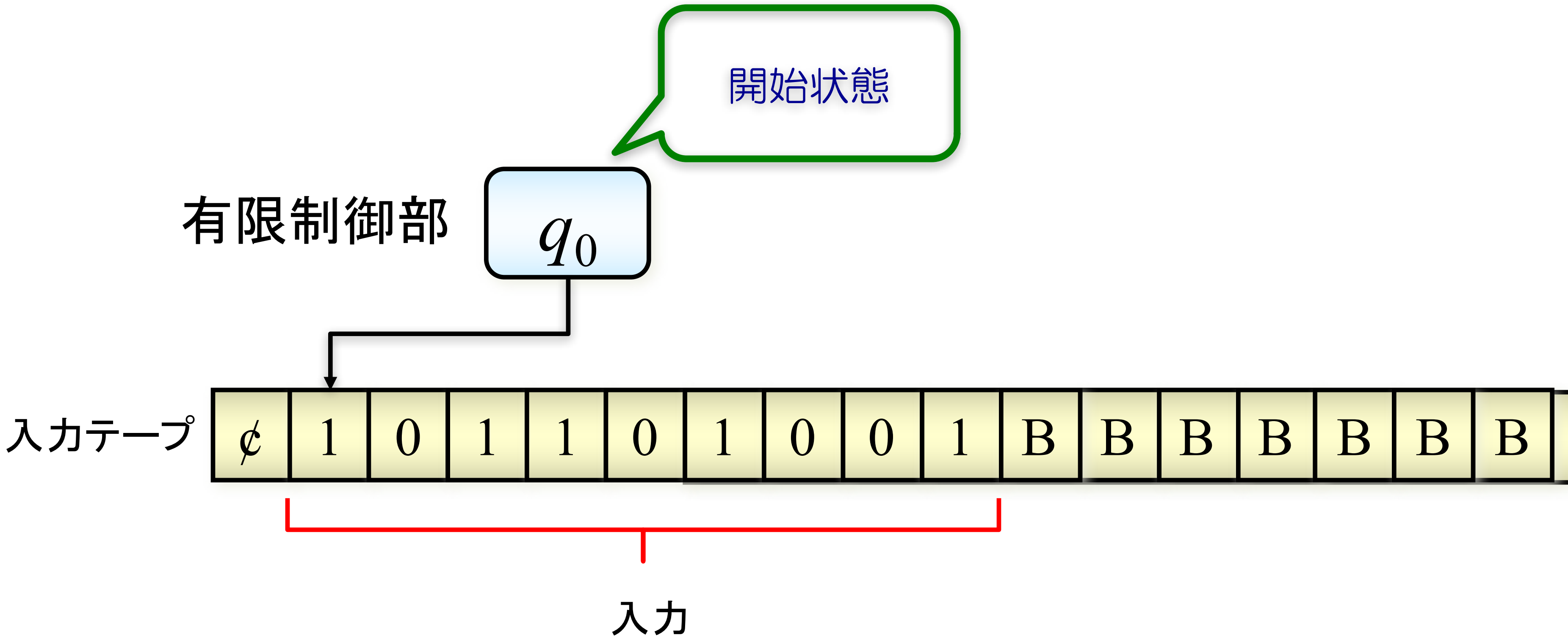
Turing機械

Turing機械とは？

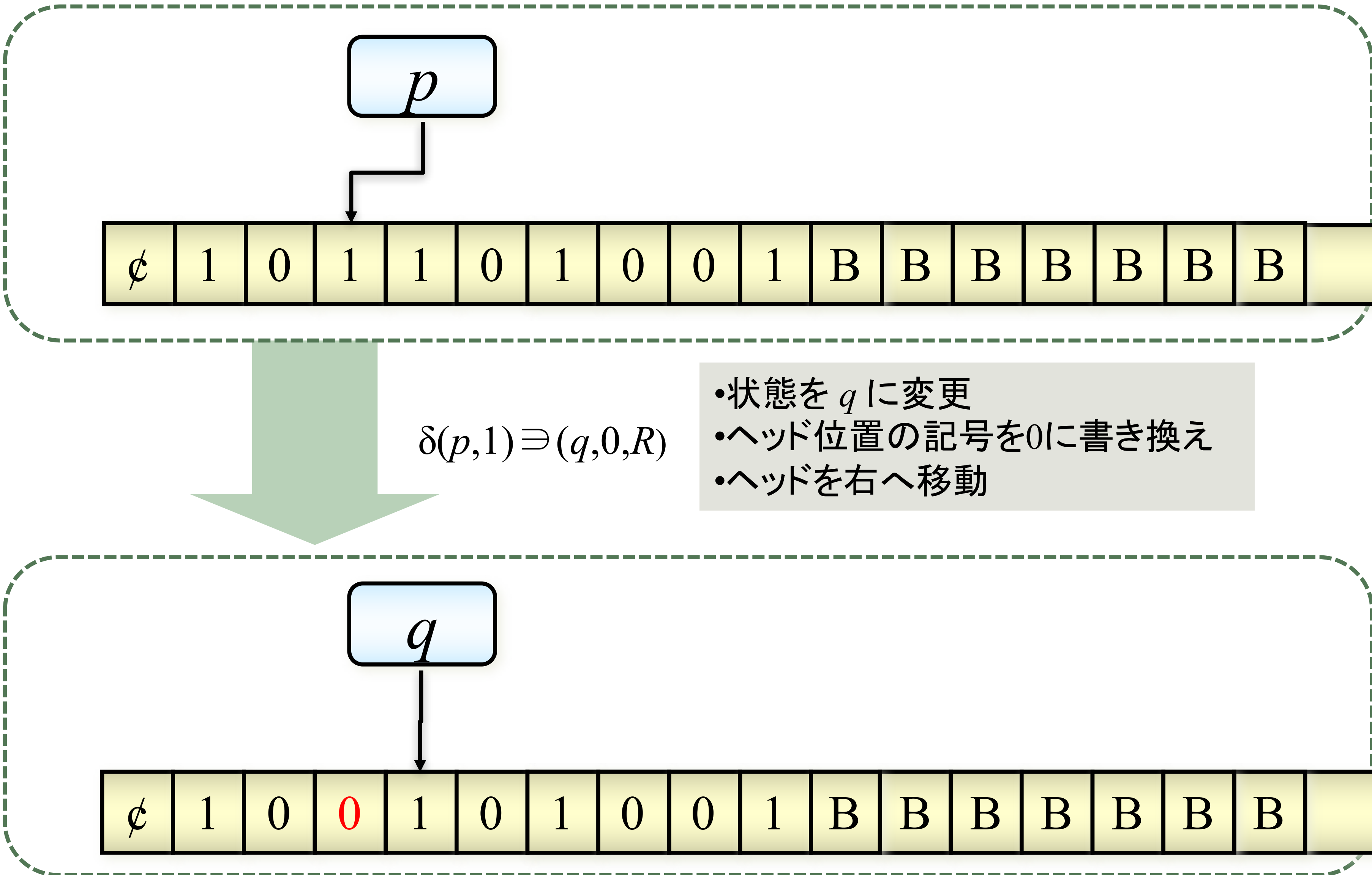
- 有限オートマトンとの違い:
  - テープヘッド位置の記号の読出しと書込みが可能である.
  - テープヘッドは左右のどちらにも動くことができる.
  - テープの長さは無限である.



計算開始時のTuring機械



Turing機械の動作の1ステップ



Turing機械の停止と入力文字列の受理

- 有限オートマトンやプッシュダウンオートマトンは，入力を読み切ったら停止するが，Turing機械のテープヘッドは左右に動くため「入力を読み切る」ということがない．
- Turing機械 M の状態を p，テープヘッド位置の記号を a とするとき，$\delta(p, a)=\varnothing$ ならば M は停止する．
  - 永遠に停止しないことも！
- Turing機械 M が停止したとき受理状態であれば，M は w を受理するという．

非決定性Turing機械の形式的定義

【定義】非決定性Turing機械は次のような6つ組
$M = (Q, \Sigma, \Gamma, \delta, q_0, F)$ である.

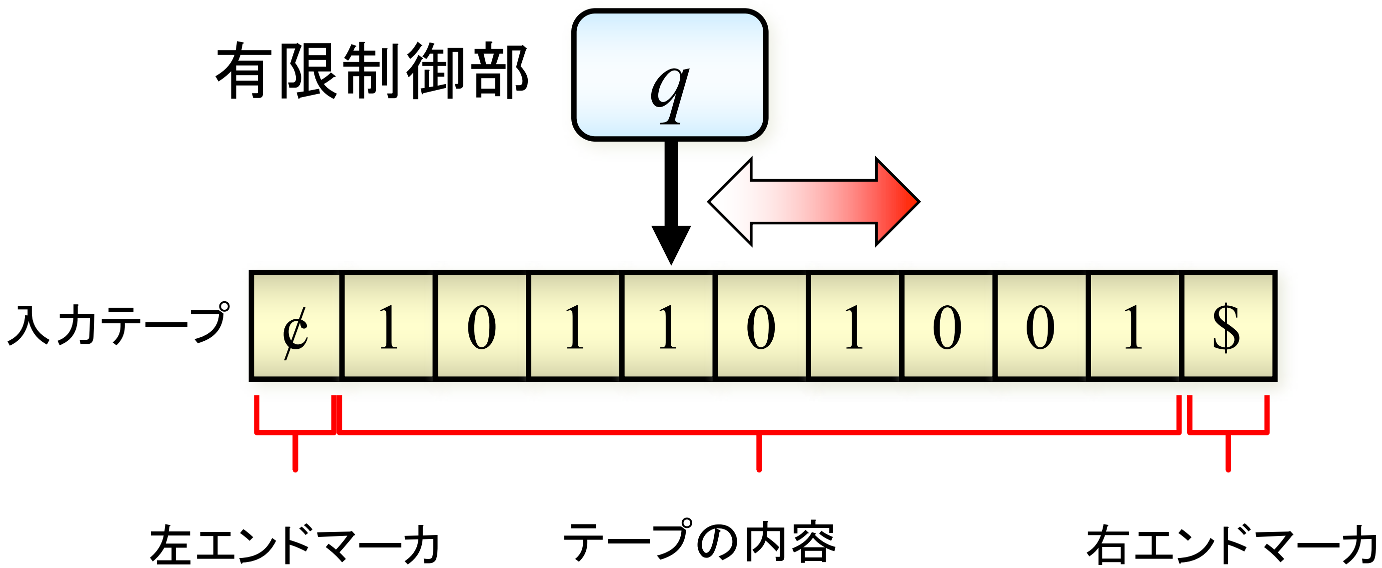
- Q は状態の空でない有限集合.
- Σ は記号の空でない有限集合で, $\Sigma \cap \{B, ¢\} = \varnothing$.
- Γ は記号の空でない有限集合で, $\Sigma \cup \{B, ¢\} \subseteq \Gamma$.
- $\delta: Q \times \Gamma \to \mathcal{P}(Q \times \Gamma \times \{\mathrm{L}, \mathrm{R}\})$は次を満たす状態遷移関数:
  任意の $p \in Q$ に対して,
  - $\delta(p, a) \subseteq Q \times (\Gamma - \{¢\}) \times \{\mathrm{L}, \mathrm{R}\} \quad (a \in \Gamma - \{¢\})$
  - $\delta(p, ¢) \subseteq Q \times \{¢\} \times \{\mathrm{R}\}$
- $q_0 \in Q$ は開始状態.
- $F \subseteq Q$ は受理状態の集合.

$$\mathcal{P}(S) = 2^S = \{X \mid X \subseteq S\}$$

線形有界オートマトン

線形有界オートマトンとは？

- Turing機械との違い:
  - テープの左端と右端に，それぞれ，$¢$, $\$$ が置かれ，入力はそれらに挟まれて与えられる．
  - テープヘッドは，テープの左端（右端）を越えて左へ（右へ）動くことはできない．



線形有界オートマトンの形式的定義

【定義】 線形有界オートマトンは次を満たす非決定性 Turing機械 $M = (Q, \Sigma, \Gamma, \delta, q_0, F)$ である.

- $\Sigma \cap \{¢, \$\} = \varnothing$, $\Sigma \cup \{¢, \$\} \subseteq \Gamma$.
- δ: $Q \times \Gamma \to Q \times \Gamma \times \{\mathrm{L}, \mathrm{R}\}$ は次を満たす状態遷移関数 : 任意の $p \in Q$ に対して,
  - $\delta(p, a) \subseteq Q \times (\Gamma - \{¢, \$\}) \times \{\mathrm{L}, \mathrm{R}\}$　$(a \in \Gamma - \{¢, \$\})$
  - $\delta(p, ¢) \subseteq Q \times \{¢\} \times \{\mathrm{R}\}$
  - $\delta(p, \$) \subseteq Q \times \{\$\} \times \{\mathrm{L}\}$

各言語族の閉性

閉性

- 和集合，連接，Kleene閉包，積集合，補集合の各演算に関する閉性は，次のようになる．

言語族	和集合	連接	Kleene閉包	積集合	補集合
句構造言語族	○	○	○	○	×
文脈依存言語族	○	○	○	○	○
文脈自由言語族	○	○	○	×	×
正規言語族	○	○	○	○	○