数理解析研究所講究録
第381巻 1980年 85-99

万能文法と言語の代数的表現

東北大　通研　　広瀬貞樹
那須正和

万能文法の概念は，笠井[1]によって導入された。笠井[1]は，最左導出万能文脈自由文法（正規言語で導出をコントロールして，その正規言語をいろいろ変えることによって，すべての文脈自由言語を最左導出で生成することができるひとつの文脈自由文法）が存在することを示した。その後，万能文法は，Rozenberg[2]，Greibach[3]，広瀬，那須[9]によって，さらに研究されている。

また，形式言語理論において，言語を代数的な演算を用いて表現する方法がいろいろ考えられている。なかでも，文脈自由言語のそのような表現に関するChomsky[4]，Stanley[5]によるつぎの結果は良く知られている。

Σを任意のアルファベットとする。あるアルファベットΣ'と，Σ'上のDyck言語D，準同形写像$h:(\Sigma')^* \longrightarrow \Sigma^*$が存在して，$\Sigma$上の任意の文脈自由言語$L$に対して，

$$L = h(D \cap R)$$

となるようなΣ'上の正規言語Rが存在する。

本報告では，言語のクラスに対して，上と同様な代数的表現が存在することと，その言語のクラスに対して最左導出万能な文脈自由文法が存在することとは，等価となることを示す。また，この結果の応用として，句構造言語や線型言語に対しても，文脈自由言語のChomsky, Stanleyの表現と同様な代数的表現が存在することを示す。

$G = \langle V_N, \Sigma, P, S \rangle$を文脈自由文法とする。$\pi : A \longrightarrow \beta$ $(A \in V_N,\ \beta \in (V_N \cup \Sigma)^*)$が，文法$G$の生成規則の集合$P$の元であるとき，任意の$\gamma \in \Sigma^*$と$\delta \in (V_N \cup \Sigma)^*$に対して，

$$\gamma A \delta \xRightarrow[lm]{\pi}{}_{G} \gamma \beta \delta$$

と書く。$\alpha_0, \alpha_1, \cdots, \alpha_n \in (V_N \cup \Sigma)^*$とし，$\pi_1, \pi_2, \cdots, \pi_n \in P$とする。各$i = 1, 2, \cdots, n$に対して，$\alpha_{i-1} \xRightarrow[lm]{\pi_i}{}_{G} \alpha_i$であるとき，

$$\alpha_0 \xRightarrow[lm]{\pi_1 \pi_2 \cdots \pi_n}{}_{G} \alpha_n$$

と書く。また，文法Gの生成規則の集合P上の言語Cに対して，

$$L_C^l(G) = \{ x \in \Sigma^* \mid {}^{\exists}\tau \in C,\ S \xRightarrow[lm]{\tau}{}_{G} x \}$$

とする。

〔定義1〕　Σを任意のアルファベットとする。GをΣ上の文脈自由文法とし，$\mathcal{L}$と$\mathcal{L}'$を任意の言語のクラスとする。$\mathcal{L}$に属するΣ上の任意の言語Kに対して，$\mathcal{L}'$に属するある言語C_Kが存在して，$L^{\ell}_{C_K}(G)=K$が成り立つとき，Gを，$\mathcal{L}'$コントロールによって，$\mathcal{L}$に対して最左導出万能なΣ上の文脈自由文法という。

〔定理1〕　$\mathcal{L}$を任意の言語のクラス，$\mathcal{L}'$を準同形写像と逆準同形写像について閉じている言語のクラスとする。Σを任意のアルファベットとするとき，つぎの (i) と (ii) とは等価である。

(i)　$\mathcal{L}'$コントロールによって，$\mathcal{L}$に対して最左導出万能なΣ上の文脈自由文法が存在する。

(ii)　アルファベットΣ'と，Σ'上の文脈自由言語L，準同形写像$h:(\Sigma')^* \longrightarrow \Sigma^*$が存在して，$\mathcal{L}$に属する$\Sigma$上の任意の言語$L_1$に対して，

$$L_1 = h(L \cap L')$$

となるような$\mathcal{L}'$に属するΣ'上の言語L'が存在する。

（証明）　(i)$\Longrightarrow$(ii) の証明

$\mathcal{L}'$コントロールによって，$\mathcal{L}$に対して最左導出万能なΣ上の文脈自由文法を，$G_0=\langle V_N, \Sigma, P_0, S\rangle$とする。文法$G_0$

から，つぎのような文脈自由文法Gを構成する。

$$G = < V_N ,\ \Sigma' ,\ P ,\ S >$$

とする。ただし，$\Sigma' = \Sigma \cup P_0$ とし，P は P_0 からつぎのように構成されるものとする。

P_0 の任意の元 $\pi_i : A \longrightarrow \alpha$ $(A \in V_N, \alpha \in (V_N \cup \Sigma)^*)$ に対して，

$$\pi_i' : A \longrightarrow \pi_i \alpha$$

を P の元とする。

g_1 を，P_0 の任意の元 $\pi_i : A \longrightarrow \alpha$ $(A \in V_N, \alpha \in (V_N \cup \Sigma)^*)$ に対しては，

$$g_1(\pi_i) = \pi_i$$

で，Σ の任意の元 a に対しては，

$$g_1(a) = \lambda$$

で定められるような，$(P_0 \cup \Sigma)^*$ から P_0^* への準同形写像とする。

また，h を，P_0 の任意の元 π_i に対しては，

$$h(\pi_i) = \lambda$$

で，Σ の任意の元 a に対しては，

$$h(a) = a$$

で定められるような $(\Sigma')^*$ から Σ^* への準同形写像とする。

このとき，つぎの命題 (*) が成立することを示す。

<u>命題(*)</u>　$\mathcal{L}'$に属するP_0上の任意の言語Cに対して，

$$L^{\ell}_{C}(G) = h(L(G) \cap g_1^{-1}(C))$$

となる。

命題(*)が成立すれば，$\mathcal{L}'$が逆準同形写像について閉じている言語のクラスであったから，$g_1^{-1}(C)$もまた$\mathcal{L}'$の元となり(i)$\Longrightarrow$(ii)がいえる。

命題(*)が成り立つことを示そう。

まず，つぎの(1)が成立することは容易に示される。

(1)　すべての$n \geq 1$と，$\pi_1, \pi_2, \cdots, \pi_n \in P_0$と，$w \in \Sigma^*$に対して，

$$S \underset{\ell m}{\overset{\pi_1\pi_2\cdots\pi_n}{\Longrightarrow}}_{G_0} w$$

であることと，$w = w_1 w_2 \cdots w_n$となるような$w_1, w_2, \cdots, w_n \in \Sigma^*$に対して，

$$S \underset{\ell m}{\overset{*}{\Longrightarrow}}_{G} \pi_1 w_1 \pi_2 w_2 \cdots \pi_n w_n$$

となることとは等価である。

さて，$w \in L^{\ell}_{C}(G_0)$とする。このとき，Cのある元$\tau = \pi_1 \pi_2 \cdots \pi_n$ ($\pi_1, \pi_2, \cdots, \pi_n \in P_0$)が存在して，

$$S \underset{\ell m}{\overset{\pi_1\pi_2\cdots\pi_n}{\Longrightarrow}}_{G_0} w$$

となる。(1)より，

$$S \underset{\ell m}{\overset{*}{\Longrightarrow}}_{G} \pi_1 w_1 \pi_2 w_2 \cdots \pi_n w_n$$

$$(\pi_1, \pi_2, \cdots, \pi_n \in P_0,\ w_1, w_2, \cdots, w_n \in \Sigma^*,\ w = w_1 w_2 \cdots w_n)$$

となるような導出が存在することがいえる。したがって，

$$\pi_1 w_1 \pi_2 w_2 \cdots \pi_n w_n \in L(G)$$

である。また，$\tau = \pi_1 \pi_2 \cdots \pi_n$ は C の元であるから，

$$\pi_1 w_1 \pi_2 w_2 \cdots \pi_n w_n \in g_1^{-1}(C)$$

である。

$$h(\pi_1 w_1 \pi_2 w_2 \cdots \pi_n w_n) = w_1 w_2 \cdots w_n = w$$

であるから，

$$w \in h(L(G) \cap g_1^{-1}(C))$$

であることがいえる。

逆に，$w \in h(L(G) \cap g_1^{-1}(C))$ とする。すると $h(w') = w$ となるような $w' \in L(G) \cap g_1^{-1}(C)$ が存在する。$w' \in g_1^{-1}(C)$ より，

$$w' = \pi_{i_1} w_1 \pi_{i_2} w_2 \cdots \pi_{i_n} w_n$$

$$(\pi_{i_1}, \pi_{i_2}, \cdots, \pi_{i_n} \in P_0,\ w_1, w_2, \cdots, w_n \in \Sigma^*,\ \pi_{i_1} \pi_{i_2} \cdots \pi_{i_n} \in C)$$

と書ける。さらに，$w' \in L(G)$ であるから (1) によって，

$$S \xRightarrow[lm]{\pi_{i_1} \pi_{i_2} \cdots \pi_{i_n}}_{G_0} w_1 w_2 \cdots w_n = w$$

となり，

$$w \in L_C^l(G_0)$$

がいえる。

以上によって命題 (*) がいえて，(i)$\Longrightarrow$(ii) が示された。

(ii)$\Longrightarrow$(i) の証明

(ii) の中の Σ' 上の文脈自由言語 L に対して，文法 $G = \langle V_N,$

Σ', P, S > を，$L=L(G)$ となるようなGreibach標準形文法とする。

g を，P の任意の元 $\pi: A \longrightarrow a\gamma$ ($A \in V_N$, $a \in \Sigma'$, $\gamma \in V_N^*$) に対して，

$$g(\pi)=a$$

で定められる P^* から $(\Sigma')^*$ への準同形写像とする。文法 G が Greibach標準形文法であることから，つぎの (2) が成り立つことが容易にわかる。

(2) P の元の系列 $\tau=\pi_1\pi_2\cdots\pi_n$ ($\pi_1, \pi_2, \cdots \pi_n \in P$) に対して，

$$S \xRightarrow[lm]{\tau}_G w \quad (w \in (\Sigma')^*)$$

となるとき，

$$g(\tau)=w$$

となる。

文法 $G=\langle V_N, \Sigma', P, S\rangle$ からつぎのような新しい文脈自由文法 G_0 を構成する。

$$G_0=\langle V_N, \Sigma, P_0, S\rangle$$

とする。ただし，P_0 はつぎのような生成規則の集合である。
P の任意の元 $A \longrightarrow a\gamma$ ($A \in V_N$, $a \in \Sigma'$, $\gamma \in V_N^*$) に対して，

$$A \longrightarrow h(a)\gamma$$

を P_0 の元とする。

また，H を，P の任意の元 $\pi: A \longrightarrow a\gamma$ ($A \in V_N$, $a \in \Sigma'$,

$r \in V_N^*$) に対して,

$$H(\pi) = \{A \longrightarrow h(a)r\}$$

で定められる P^* から P_0^* への準同形写像とする。このとき容易につぎの (3) がいえる。

(3) $S \xrightarrow[\ell m]{\tau}_G w \quad (w \in (\Sigma')^*)$

$S \xrightarrow[\ell m]{H(\tau)}_{G_0} w_0 \quad (w_0 \in \Sigma^*)$

であれば,

$$h(w) = w_0$$

である。

つぎの命題 (*) が成立することを示す。

<u>命題 (*)</u> $\mathcal{L}'$ に属する Σ' 上の任意の言語 L' に対して,

$$h(L \cap L') = L^{\ell}_{H(g^{-1}(L'))}(G_0)$$

となる。

命題 (*) が成立すれば, $\mathcal{L}'$ が準同形写像と逆準同形写像について閉じている言語のクラスであるので, $H(g^{-1}(L'))$ もまた $\mathcal{L}'$ の元となり, (ii) $\Longrightarrow$ (i) が成り立つ。

上の命題 (*) が成り立つことを示そう。

$w \in h(L \cap L')$ とする。$h(w') = w$ となる $w' \in L \cap L'$ が存在する。$w' \in L$ であるから P の元の系列 τ が存在して,

$$S \xrightarrow[\ell m]{\tau}_G w'$$

となる。$S \xrightarrow[\ell m]{\tau}_G w'$ であるから, (2) より $g(\tau) = w'$ であり,

$$S \xRightarrow[em]{H(\tau)}_{G_0} h(w') = w$$

である。$g(\tau) = w'$, $w' \in L'$ であるから，

$$\tau \in g^{-1}(w') \subseteq g^{-1}(L')$$

である。したがって

$$H(\tau) \in H(g^{-1}(L'))$$

であることがいえて，

$$w \in L^{\ell}_{H(g^{-1}(L'))}(G_0)$$

となる。

逆に，$w \in L^{\ell}_{H(g^{-1}(L'))}(G_0)$ とする。$S \xRightarrow[em]{\tau_0}_{G_0} w$ となるような $\tau_0 \in H(g^{-1}(L'))$ が存在する。$\tau_0 \in H(g^{-1}(L'))$ であるから，$H(\tau) = \tau_0$ となるような $\tau \in g^{-1}(L')$ が存在する。w' を，

$$S \xRightarrow[em]{\tau}_{G} w'$$

となるものとする。$H(\tau) = \tau_0$ であるから，H の定義から明らかなように，τ の中の規則と τ_0 の中の規則で対応するものは，それぞれ終端記号については異なっているが，変数については全く同じであるから，$w' \in (\Sigma')^*$ であることがいえ，$w' \in L$ となる。また，(2)から $g(\tau) = w'$ を得る。したがって，

$$w' = g(\tau) \in g\,g^{-1}(L') = L'$$

であることがいえ，$w' \in L \cap L'$ となる。(3)から $h(w') = w$ であることがいえるので，

$$w \in h(L \cap L')$$

を得る。

以上によって，命題 (＊) がいえて，(ii) $\Longrightarrow$ (i) が示された。

〔証明終り〕

さて，定理1の応用をいくつか述べよう。

Dyck言語 D を用いて，任意の文脈自由言語を表現するつぎのような結果が，Chomsky[4]，Stanley[5] によって示された。

〔定理2〕　(Chomsky，Stanley)

Σ を任意のアルファベットとする。あるアルファベット Σ' と，Σ' 上のDyck言語 D，準同形写像 $h:(\Sigma')^* \longrightarrow \Sigma^*$ が存在して，Σ 上の任意の文脈自由言語 L に対して，

$$L = h(D \cap R)$$

となるような Σ' 上の正規言語 R が存在する。

定理1と定理2を用いると，笠井[1] によって示された，最左導出万能文脈自由文法が存在するという結果の別証明が得られる。

また，句構造言語と線型言語に対しても，Chomsky，Stanley の文脈自由言語に対する表現と同じような代数的表現が存在することがいえる。

〔定理3〕　(Greibach[3])

Σ を任意のアルファベットとする。「線型言語のクラス」

コントロールによって，「句構造言語のクラス」に対して最左導出万能なΣ上の文脈自由文法が存在する。

定理1と定理3を用いて，つぎの定理がただちに得られる。

〔定理4〕　Σを任意のアルファベットとする。あるアルファベットΣ'と，Σ'上の文脈自由言語L，準同形写像h：$(\Sigma')^* \longrightarrow \Sigma^*$が存在して，$\Sigma$上の任意の句構造言語$L_1$に対して，

$$L_1 = h(L \cap L')$$

となるようなΣ'上の線型言語L'が存在する。

（証明）線型言語のクラスが，準同形写像と逆準同形写像について閉じているので，定理1と定理3から明らかである。

〔証明終り〕

さらに，つぎの定理が得られる。

〔定理5〕　Σを任意のアルファベットとする。つぎの，(i),(ii)の双方を満足するような，アルファベットΣ'，Σ'上のDyck言語D，準同形写像h：$(\Sigma')^* \longrightarrow \Sigma^*$が存在する。

(i)　Σ上の任意の句構造言語L_1に対して，

$$L_1 = h(D \cap L)$$

となるようなΣ'上の線型言語Lが存在する。

(ii)　Σ上の任意の文脈自由言語L_2に対して，

$$L_2 = h(D \cap R)$$

となるようなΣ'上の正規言語Rが存在する。

（証明）定理1の証明と，定理3より，あるアルファベットP_0と，$\Delta = \Sigma \cup P_0$上の文脈自由言語L_1'と，準同形写像h_1: $\Delta^* \longrightarrow \Sigma^*$（$\Sigma$の任意の元$a$に対しては，$h_1(a) = a$，$P_0$の任意の元$\pi$に対しては，$h_1(\pi) = \lambda$となる。）が存在して，$\Sigma$上の任意の句構造言語$L_1$に対して，

$$L_1 = h_1(L_1' \cap L_1'')$$

となるようなΔ上の線型言語L_1''が存在することがいえる。

また，定理2より，あるアルファベットΣ'と，Σ'上のDyck言語D，準同形写像$h_2 : (\Sigma')^* \longrightarrow \Delta^*$が存在して，$\Delta$上の任意の文脈自由言語は，$\Sigma'$上の適当な正規言語$R'$によって，$h_2(D \cap R')$の形に書ける。したがって，

$$L_1' = h_2(D \cap R')$$

となるようなΣ'上の正規言語R'が存在する。よって，

$$\begin{aligned} L_1 &= h_1(h_2(D \cap R') \cap L_1'') \\ &= h_1 h_2(D \cap R' \cap h_2^{-1}(L_1'')) \end{aligned}$$

となる。$h = h_1 h_2$，$L = R' \cap h_2^{-1}(L_1'')$とおく。線型言語のクラスは，逆準同形写像と，正規言語との共通集合をとる演算について閉じているから，Lは線型言語であり，

$$L_1 = h(D \cap L)$$

となる。

さらに，Σ上の任意の文脈自由言語L_2に対して，

$$L_2 = h(D \cap R)$$

となるようなΣ'上の正規言語Rが存在することを示す。L_2は，Σ上の言語であるから，$\Delta = \Sigma \cup P_0$上の言語でもある。したがって，Σ'上の正規言語Rが存在して，

$$L_2 = h_2(D \cap R)$$

となる。また，準同形写像$h_1 : \Delta^* \longrightarrow \Sigma^*$は，$\Sigma$の任意の元$a$に対しては，$h_1(a) = a$，$P_0$の任意の元$\pi$に対しては，$h_1(\pi) = \lambda$となる準同形写像であるから，

$$L_2 = h_1(L_2)$$

が成り立つ。したがって，

$$L_2 = h_1(L_2) = h_1 h_2(D \cap R) = h(D \cap R)$$

となる。

〔証明終り〕

Greibach[(3)]によるつぎの結果と，定理1とから，ただちに線型言語に対してもChomsky, Stanleyの文脈自由言語に対する表現と同様な代数的表現が存在することがいえる。

〔定理6〕　(Greibach[(3)])

Σを任意のアルファベットとする。「正規言語のクラス」コントロールによって，「線型言語のクラス」に対して，最

左導出万能なΣ上の線型文法が存在する。

〔定理7〕 Σを任意のアルファベットとする。つぎの(i)を満足するような，アルファベットΣ'，Σ'上の線型言語L，準同形写像$h:(\Sigma')^* \longrightarrow \Sigma^*$が存在する。

(i) Σ上の任意の言語L_3に対して，L_3が線型言語ならば，

$$L_3 = h(L \cap R)$$

となるようなΣ'上の正規言語Rが存在し，またその逆も成り立つ。

（証明）定理1の証明と，定理6とから，明らかに，アルファベットΣ'と，Σ'上の線型言語L，準同型写像$h:(\Sigma')^* \longrightarrow \Sigma^*$が存在して，$\Sigma$上の任意の線型言語$L_3$に対して，

$$L_3 = h(L \cap R)$$

となるようなΣ'上の正規言語Rが存在することがわかる。

また，線型言語のクラスが，正規言語との共通集合をとる演算と，準同形写像について閉じているから，任意の線型言語L，正規言語R，準同形写像hに対して，

$$L_3 = h(L \cap R)$$

で表わされる言語L_3は，線型言語である。

〔証明終り〕

[References]

(1) Kasai,T. (1975), A universal context-free grammar, Inform. Contr. 28, 30-34.

(2) Rozenberg,G. (1977), A note on universal grammars, Inform. Contr. 34, 172-175.

(3) Greibach,S.A. (1978), Comments on universal and left universal grammars, context-sensitive languages, and context-free grammar forms, Inform. Contr. 39, 135-142.

(4) Chomsky,N. (1962), Context-free grammars and pushdown storage, M.I.T. Res. Lab. Electron. Quart. Prog. Report. 65.

(5) Stanley,R.J. (1965), Finite state representation of context-free languages, M.I.T. Res. Lab. Electron. Quart. Prog. Report. 76, 276-279.

(6) Ginsburg,S. (1966), "The Mathematical Theory of Context-Free Languages", McGraw Hill, New York.

(7) Hopcroft,J.E.,Ullman,J.D. (1969), "Formal Languages and Their Relations to Automata", Addison-Wesley, Reading, Mass.

(8) Harrison,M.A. (1978), "Introduction to Formal Language Theory", Addison-Wesley, Reading, Mass.

(9) 広瀬,那須. 万能文法に関する二,三の考察,信学論,採録決定.