数理解析研究所講究録
123巻 1971年 103-113

String Controlled Grammar

名古屋大学　工学部　伊藤英則
稲垣康善
福村晃夫

§1. はじめに

生成規則の適用順序を制御する文法が最近いろいろ考えられている。これらは異った制御の方法を用いているが，いずれも複雑さの低いクラスの文法に制御を加えることにより高いクラスの言語を生成できる特徴を持っている。

そこで一つの新しい試みとして，CF文法の文法規則の適用順序を補助記号の系列によって制御する String Grammar を考えた。これは core 部と String 部を有していて，core 部にはCF文法の生成規則を持ち，String 部には補助記号の系列を持った文法である。String 部の補助記号の用い方によって i) nested stack 形 (ns)，ii) stack 形 (s)，iii) pushdown 形 (pd) String Grammar (SG) に分類する。

本報告では，SG が生成する言語の族のレベルについて述べる。まず，core 部が ε-free CF 形である，pd 形 SL は ε-free CFL の族を真に含み，ns 形 SL は CSL の族と一

致することを明らかにする。つぎに，εを許したCF形の生成規則であるならば，pd形SLはrecursively enumerable setの族になることを示す。さらにSLの族はAFLになることを述べる。

§2. 諸定義

〔定義1〕 ns形SGを$G_n=(V,\Sigma,\Gamma,P,\sigma,Z)$とする。i) V；記号の有限集合，ii) Σ；終端記号の有限集合，iii) Γ；補助記号の有限集合，iv) P；$(V-\Sigma)\times\Gamma\times V^+\times\Gamma^*\times D$の部分集合で生成規則の集合を示す。$(\xi,A,u,\alpha,i)\in P$を$(\xi,A)\to(u,\alpha,i)$と記す。ここに，$\xi\in V-\Sigma$，$A\in\Gamma$，$u\in V^+$，$\alpha\in\Gamma^*$，$i\in D=\{-1,0,1\}$である。さらに，$(\xi,A)\longrightarrow(u,\varepsilon,i)$のとき，$u\neq V-\Sigma$であり，$A$はstring部の最左端の補助記号である。v) $\sigma\in V-\Sigma$；core部の初期記号，vi) $Z\in\Gamma$；string部の初期記号である。

〔定義2〕 S形SGを$G_s=(V,\Sigma,\Gamma,P,\sigma,Z)$とする。$G_s$は生成規則がつぎのように制限されたns形SGである。すなわち，$P=P_1\cup P_2$，$P_1\subseteq(V-\Sigma)\times\Gamma\times V^+\times\Gamma^*$，$P_2\subseteq(V-\Sigma)\times\Gamma\times V^+\times D$，$D=\{-1,0,1\}$。ただし，$P_1$の生成規則は最左端にある補助記号に適用されるのみである。

〔定義3〕 pd形SGを$G_p=(V,\Sigma,\Gamma,P,\sigma,Z)$と

する。G_pはPとして定義2のP_1の形のもののみをもつS形SGである。

〔定義4〕SG, $G=(V, \Sigma, \Gamma, P, \sigma, Z)$において，$x$, $y \in V^*$, $\xi \in V-\Sigma$, $A \in \Gamma$, $\alpha, \beta, \gamma, \delta \in \Gamma^*$, $u \in V^+$なるとき，もし，$(\xi; A) \longrightarrow (u, \beta, i)$がPに含まれていれば$(x\xi y, \alpha \uparrow A \gamma) \Longrightarrow (xwy, \delta)$なる関係$\Longrightarrow$が成立つ。ただし，$i=-1$ならば，$\delta = A_0 \cdots A_{j-1} \uparrow A_j \beta \gamma$, $i=0$ならば，$\delta = \alpha \uparrow \beta \gamma$, $i=+1$ならば，$\delta = \alpha B_0 \uparrow B_1 \cdots B_j \gamma$，さらに$\beta = \varepsilon$, $i=+1$ならば，$\delta = \alpha C_0 \uparrow C_1 \cdots C_j$。ここに，$\alpha = A_0 \cdots A_j$, $\beta = B_0 \cdots B_j$, $\gamma = C_0 \cdots C_j$，また，$\Longrightarrow$のreflective, transitive closureを$\overset{*}{\Longrightarrow}$とする。

〔定義5〕SG, $G=(V, \Sigma, \Gamma, P, \sigma, Z)$によって生成される言語をL(G)とすれば，$L(G)=\{w \mid (\sigma, \uparrow Z) \overset{*}{\Longrightarrow} (w, \uparrow Z), w \in \Sigma^*\}$. nS形SG, G_nによって生成された言語$L(G_n)$をns形SL (String language), G_sによって生成された言語$L(G_s)$をS形SL，pd形SG, G_pによって生成された言語$L(G_p)$をpd形SLという。また，$L(G_n)$, $L(G_s)$, $L(G_p)$の族をそれぞれ，$\mathcal{L}_n$, $\mathcal{L}_s$, $\mathcal{L}_p$と記す．定義．1, 2, 3より$\mathcal{L}_p \subseteq \mathcal{L}_s \subseteq \mathcal{L}_n$が成立つ。

〔定義6〕SG, $G=(V, \Sigma, \Gamma, P, \sigma, Z)$において，生成規則のすべての元がつぎの，i), ii)の条件を満しているとき，

normal SGという。$(\xi, A) \to (u, \alpha, i)$ において，

i) $1 \leq |u| \leq 2$， ii) $0 \leq |\alpha| \leq 2$．とくに区別するときは $G_0 = (V_0, \Sigma_0, \Gamma_0, P_0, \sigma_0, Z_0)$ と記す。ここに，$|w|$ は語 w の長さを表わす。

〔定理1〕任意のSG，$G = (V, \Sigma, \Gamma, P, \sigma, Z)$ が生成する言語を $L(G)$ とする。このとき，$L(G) = L(G_0)$ である normal SG，G_0 が存在する。（証明略）

§3. ns形SGの性質

〔例1〕 ns形SG，$G_n = (V, \Sigma, \Gamma, P, \sigma, Z)$ において

$$
\begin{aligned}
P = \{ & (\sigma, Z) \longrightarrow (\xi, Z'AZ, 1),\ (\xi, A) \to (\xi, AA', 1) \\
& (\xi, A') \longrightarrow (\xi, A, 1),\quad (\xi, Z) \longrightarrow (\eta, Z, -1) \\
& (\eta, A) \longrightarrow (\eta, A, -1),\quad (\eta, Z') \longrightarrow (\xi, Z', 1) \\
& (\eta, Z') \longrightarrow (b\zeta, \varepsilon, 0),\quad (\zeta, A) \longrightarrow (a\zeta, \varepsilon, 1) \\
& (\zeta, A) \to (a, \varepsilon, 0) \}
\end{aligned}
$$

とすると， $L(G_n) = \{ b a^{2^n} \mid n \geq 1 \}$

〔補題1〕 ns形SG，$G_n = (V, \Sigma, \Gamma, P, \sigma, Z)$ が生成する言語 $L(G_n)$ を認識可能な nondeterministic lba A が存在する。

（証明）任意のns形SG，G_n が生成する言語 $L(G_n)$ の任意の元 w に対して，w を生成する G_n の導出 D が必ずひとつは存在している。そこで，D；$(\sigma, \uparrow Z) = (w_0, \alpha_0 \uparrow \beta_0 Z) \Longrightarrow$

$(w_1, \alpha_1 \uparrow \beta_1 z) \Longrightarrow \cdots \Longrightarrow (w_r, \alpha_r \uparrow \beta_r z) = (w, \uparrow z)$ とすれば，core部は ε-free の生成規則によって導出されていることより，$1 = |w_0| \leq |w_1| \leq |w_2| \leq \cdots \leq |w_r| = |w| = n$
また，$L(G)$の定義より，wは$(\sigma, \uparrow z) \xRightarrow{*} (w, \uparrow z)$によって導出されるので，導出Dが完了するまでには，string部のzの左側に書込れた補助記号系列はすべて消去されなければならない。しかも，ns形SGの定義からPの元，$(\xi, A) \longrightarrow (\eta, \varepsilon, i)$，$\xi, \eta \in V - \Sigma$，$A \in \Gamma$，$i \in D$ なる形の生成規則は許されていない。したがって，string部を消去するためには$(\xi, A) \rightarrow (u, \varepsilon, i)$，$u \in V^{+}$ かつ $u \notin V - \Sigma$ なる形の生成規則を使用しなければならない。そこで，もしstring部の長さが $m\ (> n+1)$のものが存在していたとすると，$(\xi, A) \longrightarrow (u, \varepsilon, i)$ の形の生成規則を少くともm回使用しなければならない。ところがm回使用すれば，core部の性質により$|w| \geq m$ でなければならない。これは，$|w| = n$ としたことに矛盾する。したがって，$\max_{0 \leq i \leq r} |\alpha_i \beta_i| \leq n$ でなければならない。以上のことより，長さnの語wを生成するとき使用するstring部の長さは高々$(n+1)$でよいことから，$L(G_n)$を認識可能な nondeterministic lba A を構成することができる。
(図1参照)。nondeterministic lba A は入力 word w ($|w| = n$) に対して，working tape は $2n+1$の長さで認識できる。

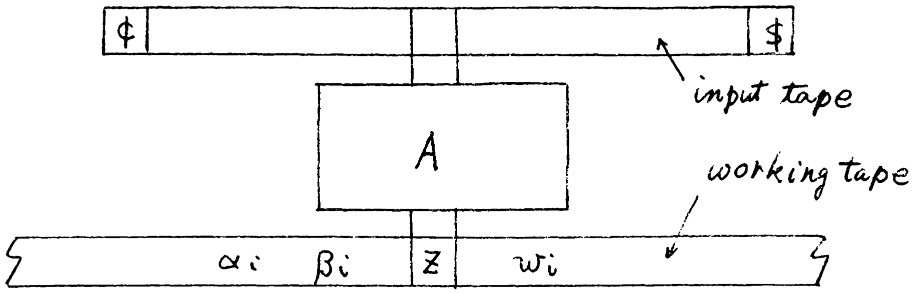

図1. lba A.

[補題2] ns形SG, G_n は任意のCSG, Gが生成する言語L(G)を生成することができる。

(証明) 任意のCSG, $G=(V, \Sigma, P, \sigma)$ は, 文献1よりつぎの3つの形の生成規則のみからなっていると考えてもよい。すなわち, $\xi, \eta, \gamma, \zeta \in V-\Sigma$, $a \in \Sigma$ に対して,

i) $\xi\eta \to \gamma\zeta$, ii) $\xi \to \gamma\zeta$, iii) $\xi \to a$

これら, i), ii), iii) の形の生成規則をns形SGによってまねることを考える。CSG, Gにおいては, i), ii) の形のみを使用して導出を行い, いったん, iii)の形の生成規則を適用したあとは, i), ii)の形の生成規則は使用しないで, 任意のword w を導出すると仮定しても一般性を失わない。ゆえに

$$\sigma = w_0 \Rightarrow w_1 \Rightarrow \cdots \Rightarrow w_l \Longrightarrow \cdots \Longrightarrow w_n = w$$

の導出において, $\sigma \overset{*}{\Rightarrow} w_l$ までは, i), ii) の形の生成規則が適用されて, $w_l \overset{*}{\Longrightarrow} w_n$ はiii)の形の生成規則のみが適用されて生成されたとする。そこでi), ii) の形の生成規則をつぎ

のようにして，ns 形 SG，G_n でまねる。

$$\text{i)}\ (\sigma_s, A_\xi) \longrightarrow (\eta, A_r, 1)$$

$$(\eta, A_\eta) \longrightarrow (\sigma_s, A_\xi, 0), \quad (\sigma_s \neq \eta)$$

$$\text{ii)}\ (\sigma_s, A_\xi) \longrightarrow (\sigma_s, A_r A_\xi, 0)$$

さらに，$\mathcal{P}$ の任意の元 A に対して

$$(\sigma_s, A) \longrightarrow (\sigma_s, A, i), \quad i \in D$$

また，初期ステップとして，

$$(\sigma_s, Z) \longrightarrow (\sigma_s, A_\sigma Z, 0)$$

上のように構成すれば，$\sigma \overset{*}{\Longrightarrow} w_\ell$ の導出を String 部でまねることができる．すなわち，

$$(\sigma_s, \uparrow Z) \Longrightarrow (\sigma_s, \uparrow A_\sigma Z) \overset{*}{\Longrightarrow} (\sigma_s, \overline{w_i} Z) \overset{*}{\Longrightarrow} (\sigma_s, \overline{w_\ell} Z)$$

ただし，$\overline{\xi} = A_\xi$ を表わす．

（注）もし，$w_{j-1} \Rightarrow w_j$ は CSG，G において，i)の形の生成規則が適用されているとすると，

$$(\sigma_s, \overline{w_j} Z) = (\sigma_s, \alpha \uparrow A_\xi A_\eta \beta Z) \Longrightarrow (\eta, \alpha A_r \uparrow A_\eta \beta Z)$$

$$\Longrightarrow (\sigma_s, \alpha A_r \uparrow A_\xi \beta Z) = (\sigma_s, \overline{w_j} Z)$$

つぎに，iii)の形の生成規則を ns 形 SG，G でまねる．

$$\text{iii)}\ (\sigma_s, A_\xi) \longrightarrow (a\sigma_s', \varepsilon, 0)$$

$$(\sigma_s', A_\xi) \longrightarrow (a\sigma_s', \varepsilon, 0)$$

$$(\sigma_s', A_\xi) \longrightarrow (a, \varepsilon, 0)$$

CSG の iii) の形の生成規則は CF 形だから，$w_\ell \overset{*}{\Longrightarrow} w_n$ は

iii)の形の生成規則を使用し，最左端導出によって生成されたと仮定してもよい。上のように構成すれば，String 部の左端から順に，CSG で，$(\xi \to a)$ によって ξ が a に書換えられるのに対して，SG では，Core 部へ a を生成してゆく。このとき，String 部の記号は左から順に消去され，最後に Z だけになり，Core 部には w_n が生成される。よって補題2が成立つ。

〔定理2〕 ns 形 SG と CSG とは等価である。

（証明） 補題1，2よりいえる。

§5. pd 形 SG の性質

〔定義7〕 pd 形 SG，$G_p=(V, \Sigma, \Gamma, P, \sigma, Z)$ が最左端導出によって生成する言語を $L_1(G_p)$ とする。また，$L_1(G_p)=L(G_p)$ となる $L(G_p)$ の族を $\mathcal{L}_p'$ とする。

〔例2〕 pd 形 SG，$G_p=(V, \Sigma, \Gamma, P, \sigma, Z)$ において，

$$
\begin{aligned}
P=\{&(\sigma, Z)\longrightarrow(\xi\eta, AZ, 0), (\xi, A)\longrightarrow(ab, \varepsilon, 0)\\
&(\xi, A)\longrightarrow(a\xi b, AB, 0), (\eta, B)\longrightarrow(c\eta, \varepsilon, 0)\\
&(\eta, B)\longrightarrow(c, \varepsilon, 0)\}
\end{aligned}
$$

とすると，$L(G_p)=L_1(G_p)=\{a^n b^n c^n \mid n \geq 1\}$ となる。

〔定理3〕 $\mathcal{L}_p'$ は ε-free CFL の族を真に含む。

（証明） 任意の ε-free CFG を $G'=(V', \Sigma', P', \sigma')$ とする。$(\xi \to w)\in P'$ のときかつそのときにかぎり $(\xi, Z)\longrightarrow(w, Z,$

$0) \in P$ とすれば，$L(G_p)=L(G')$ なる pd 形 SG，$G=(V,\Sigma,\Gamma,P,\sigma,Z)$ を構成できる。しかも，P の core 部が ε-free CF 形だから，$L(G_p)$ のすべての元は最左端導出によって生成されうるように構成できる。したがって，$L(G_p)=L_1(G_p)$．つぎに例2により $\{a^n b^n c^n \mid n \geq 1\}$ は CFL でない CSL であり，$\mathcal{L}_p'$ に含まれるから定理3がいえる。

〔補題3〕 pd 形 SG において，core 部に ε を許したものを，G^* とする。すなわち，$G^*=(V,\Sigma,\Gamma,P^*,\sigma,Z)$ の生成規則の集合 P^* は $(V-\Sigma)\times\Gamma\times V^*\times\Gamma^*$ の部分集合である。このとき，任意の CFG，CFL，C に対して文献2が定義する $L_C(G)$ と $L_1(G_p^*)$ とが等しい，pd 形 SG，G_p^* が存在する。

（証明）文献2と同じ記号を用いる。$L_C(G)=g_G(\{\sigma\},C\cap P^*)\cap\Sigma^*$ である。$L_C(G)$ の任意の元を w とすると，$w=g_G(\sigma,c)$，$c\in C$ が存在する。すなわち，

$$\sigma=w_0 \underset{1}{\overset{p_1}{\Longrightarrow}} w_1 \underset{1}{\overset{p_2}{\Longrightarrow}} \cdots \underset{1}{\overset{p_r}{\Longrightarrow}} w_r=w, \quad c=p_1p_2\cdots p_r$$

なる導出が存在している。ここに，$w_{i-1}\underset{1}{\overset{p_i}{\Longrightarrow}}w_i$ は，w_{i-1} から w_i の導出は，p_i の生成規則を w_{i-1} の最左端の変数に適用して生成されていることを示す。一方，$C=L(G')$ なる CFG，G' が存在する。それを，$G'=(V',\Sigma',P',\sigma')$，$P'=\{\xi'\to w'\}$ とする。そこで，pd 形 SG，$G_p^*=(V,\Sigma,P^*,\Gamma,\sigma,Z)$ を $\Gamma=V'$，$P=\{(\sigma,Z)\to(\sigma,\sigma'Z,0)\}\cup\{(\xi,\xi')\to(\xi,u',0)\mid\xi$

$\in V-\Sigma,\ \xi' \longrightarrow v' \in P' \} \cup \{(\xi, p) \longrightarrow (v, \varepsilon, 0) \mid p: \xi \longrightarrow v$ $\in P\}$ のように構成すれば，$L_c(G)$ における w に対する導出に対応して，G_s においてつぎの導出が存在する。

$$(\sigma, z) \underset{1}{\Longrightarrow} (\sigma, \sigma' z) = (w_0, w_0') \overset{*}{\underset{1}{\Longrightarrow}} (w_r, w_r') = (w, z)$$

また，この逆も成立つ．したがって，$L_c(G) = L_1(G_p^*)$ なる pd 形 SG，G_p^* が存在する．

〔補題4〕 任意の recursively enumerable set U に対して，$U = L_c(G)$ を満す，CFG，G．CFL，C とが存在する．

（証明） 文献1の定理2.1による．

〔定理5〕 任意の recursively enumerable set U に対して，$U = L_1(G_p^*)$ を満す pd 形 SG，G_p^* が存在する。

（証明） 補題 3．4による．

§6．SL の Closure property

〔補題6〕 $\mathcal{L}_p, \mathcal{L}_s, \mathcal{L}_n$ は $\cup$，$\cdot$，$+$ の演算のもとでとじている．

〔補題7〕 $\mathcal{L}_p, \mathcal{L}_s, \mathcal{L}_n$ は regular set との intersection のもとでとじている。

〔補題8〕 $\mathcal{L}_p, \mathcal{L}_s, \mathcal{L}_n$ は ε-free homomorphism，inverse homomorphism のもとでとじている．

〔定理5〕 $\mathcal{L}_p, \mathcal{L}_s, \mathcal{L}_n$ は AFL である．

（証明） 補題 5．6．7 による

§7. あとがき

今後の課題として，i) $\mathcal{L}_p$, $\mathcal{L}_s$, $\mathcal{L}_u$ との間に真の包含関係が成立つか．ii) 生成規則の適用順序が制御されている種々の既存の文法との対応関係について調らべる必要がある。

謝辞，日頃熱心に御指導賜る研究室の皆様，東北大学本多波雄教授に深謝する．

§8 文献

1) Kuroda, S.Y. "Classes of Languages and Linear-Bounded Automata" Inf & Cont. Vol 7, p207-223

2) Ginsburg, S. "Control sets on Grammars" mathematical System Theory, Vol. 2 p157-177 (1968)