Extended Right Precedence Grammars and Analyzing Technique for Them

SHOJI SEKIMOTO*

1. *Introduction*

The right precedence grammars have been originally proposed by K. Inoue, and have the advantage, as compared with precedence grammars, that they have only right precedence tables and need no left precedence tables for parsing.

In this paper, we concern with grammatical rules given by Inoue, loosen the restriction of these rules and make clear that the class of grammars given in this paper includes the class of right precedence grammars.

2. *Right Precedence Grammars*

For a given context free grammar $G=(V_N, V_T, P, S)$, we define a grammar G' as follows and if this satisfies next four conditions (i), (ii), (iii) and (iv), then we call G' to be a right precedence grammar.

$$G'=(V_N', V_T', P', S'),\ V_N'=V_N \cup \{S'\},\ V_T'=V_T \cup \{\vdash, \dashv\}$$
$$P'=P \cup \{S' \to \vdash S \dashv\},\ V'=V_N' \cup V_T'$$

(i) We can produce $A \overset{*}{\Rightarrow} U$ for $^{\forall}A \in V_N'$ and some $U \in V_T'^{*}$.

(ii) $A \to \alpha$ and $B \to \alpha$ $(A \neq B)$ do not exist together in P.

(iii) There is a unique and non-commutative right precedence relation between A and $B (A, B \in V')$ if $S \overset{*}{\Rightarrow} \alpha A B \beta$.

The right precedence relations are decided as follows:

a) $B \lessdot C$ if $A \to \beta B E \gamma$ and $C \in L_T(E)$,

b) $B \gtrdot C$ if $A \to \beta D E \gamma$, $B \in R(D)$ and $C \in L_T(E)$.

(iv) If $a \to \alpha\beta$ and $B \to \beta$ exist in P then there are not

$C \to rHi(\alpha)D$ and $D \overset{*}{\Rrightarrow} T_{n-i}(\alpha) B \delta'$,

where $n=|\alpha|$ and $n \geq i > 0$.

Note: $L(A)=\{F \in V' \mid A \overset{*}{\Rightarrow} F\alpha,\ \alpha \in V'^{*}\}$,

$R(A)=\{F \in V' \mid A \overset{*}{\Rightarrow} \alpha F,\ \alpha \in V'^{*}\}$ for $^{\forall}A \in V_N'$.

$L_T(A)=\{F \in V_T' \mid A \overset{*}{\Rrightarrow} F\alpha,\ \alpha \in V'^{*}\}$ for $^{\forall}A \in V'$,

where $\xi \overset{*}{\Rrightarrow} \eta$ means $\xi \overset{*}{\Rightarrow} \eta$ or $\xi = \eta$.

This paper first appeared in Japanese in Joho-Shori (Journal of the Information Processing Society of Japan), Vol. 12, No. 9 (1971), pp. 543~548.

* Mitsubishi Electric Corporation

$$Hi(\alpha)=\begin{cases} a_1\cdots a_i \text{ for } 1\leq i\leq n \\ \text{empty string for } i=0, \end{cases}$$

$$Ti(\alpha)=\begin{cases} a_{n-i+1}\cdots a_n & \text{for } 1\leq i\leq n \\ \text{empty string} & \text{for } i=0, \end{cases}$$

$$\text{for } {}_V\alpha=a_1\cdots a_n\in V'^*$$

Now the context free grammars with conditions (i), (ii), (iii) and (iv) are called 'simple right precedence grammars'. Before we give extended right precedence grammars, we rewrite (iv) to (iv°) as follows:

(iv°) If $A\rightarrow\alpha\beta$ and $B\rightarrow\beta$ exist in P' then

$$C_1\rightarrow\gamma_1 H_{n-i}(\alpha)D_1\varepsilon_1 \text{ and}$$

$$D_1\overset{*}{\Rightarrow}T_i(\alpha)B\delta_1$$

do not exist together for $n=|\alpha|$ and $0\leq i<n$, especially if the latter is true then

$$C_2\rightarrow\gamma_2 H_{n-i-i_1}(\alpha)D_2\varepsilon_2 \text{ and}$$

$$D_2\overset{*}{\Rightarrow}T_{i_1}(H_{n-i}(\alpha))D_1\delta_2$$

do not exist together for i_1, $0\leq i+i_1<n$.

Repeating this process for l times, for $0\leq i+i_1+\cdots+i_{l-1}<n$ and $k=i+i_1+i_2+\cdots+i_{l-1}$, if $D_l\overset{*}{\Rightarrow}T_{i_{l-1}}(H_{n-k+i_{l-1}}(\alpha))D_{l-1}\delta_l$ is true them

$$C_{l+1}\rightarrow\gamma_{l+1}H_{n-k-i_l}(\alpha)D_{l+1}\varepsilon_{l+1} \text{ and } D_{l+1}\overset{*}{\Rightarrow}T_{i_l}(H_{n-k}(\alpha))D_l\delta_{l+1}$$

do not exist together for i_l, $0\leq i+i_1+\cdots+i_{l-1}+i_l<n$. We will use the rule (iv°) instead of (iv) in section 3.

3. *An Extension of Right Precedence Grammars*

3.1 *On the Rule (iv°)*

We will give an extension of the rule (iv°), and get an extended class of the right precedence grammars by using it instead of (iv°). We will define two subsets of V_T', $Q_T(A)$ and $Q_T(\alpha|A)$, where $A\in V_N'$ and $\alpha\in V'^*$.

For a given $A(\in V_N')$, $P_0(A)$ is defined as a subset of P, where each element of $P_0(A)$ includes at least one A in its right hand side excluding the right most position of it.

If A is the start symbol then $P_0(A)=\phi$ (: empty set).

Then let $Z(A)=P_0(A)=\phi$.

Next, for A, if there exist productions such that the right most symbols of their right hand sides are A, then let the left hand side symbols ($\neq S$) of these productions be

$$P_{0i},\ i=1,2,\cdots,n_0$$

and let $P_1(A)=\bigcup_{i=1}^{n_0}P_0(P_{0i})$. If there does not exist such a P_{0i}, let $Z(A)=P_0(A)$.

Consecutively, for these P_{0i}, decide

$$P_{1i},\ i=1,2,\cdots,n_1$$

as P_{0i} for A and let $P_2(A)=\bigcup_{i=1}^{n_1} P_0(P_{1i})$.

If there is no P_{1i} for all P_{0i} then let $Z(A)=P_0(A)\cup P_1(A)$. Repeating this process, we may decide a finite number of sets $P_0(A), P_1(A), \cdots, Pm(A)$ and $Z(A)=\bigcup_{j=1}^{m} Pj(A)$, where m is determined corresponding to A, and $Pm'(A)=\phi$ for $\forall m'(>m)$. We may show that each element of the subset $P_0(A)$ in $Z(A)$ has at least one A in its right hand side and has the form

$$P_{0i}\rightarrow\gamma AW_{0k}\delta(W_{0k}\in V').$$

Then we may evaluate $L_T(W_{0k})$, $k=1,2,\cdots,n_0'$ for each W_{0k} which is the immediate right neighbor of each A in the right hand side of each production included in $P_0(A)$. Similarly we may evaluate

$L_T(W_{jk})$, $k=1,2,\cdots,n_j'$ for $Pj(A)$, $j=1,2,\cdots,m$, where W_{jk} is in V' and is the immediate right neighbor of $P_{j-1l}(1\leq l\leq n_{j-1})$ in the right hand side of some production included in $Pj(A)$.

Then we define $Q_T(A)$ by

$$Q_T(A)=\bigcup_{j=0}^{m}\bigcup_{k=1}^{nj'}(L_T(W_{jk})).$$

$Q_T(\alpha|A)$ is similarly defined by using αA instead of A in the case of $Q_T(A)$. It is clear that $Q_T(A)$ is a set of terminal symbols which have precedence relations with A as the immediate right neighbors of A. By the definition of $Q_T(A)$ and $Q_T(\alpha|A)$, the rule (iv°) is extensively rewritten as follows:

(iv′) If $A\rightarrow\alpha\beta$ and $B\rightarrow\beta$ exist in $P, Q_T(A)\cap Q_T(B)=\phi$, and $n=|\alpha|$ then

$$C_1\rightarrow\gamma_1 H_{n-i}(\alpha)D_1\varepsilon_1 \text{ and } D_1\overset{*}{\Rightarrow}T_i(\alpha)B\delta_1$$

do not exist together, and if the latter is true for $0\leq i<n$ then

$$C_2\rightarrow\gamma_2 H_{n-i-i_1}(\alpha)D_2\varepsilon_2 \text{ and } D_2\overset{*}{\Rightarrow}T_{i_1}(H_{n-i}(\alpha))D_1\delta_2$$

do not exist together for i_1, $0\leq i+i_1<n$. By repeating this process for l times, if

$$D_1\overset{*}{\Rightarrow}T_{i_{l-1}}(H_{n-k+i_{l-1}}(\alpha))D_{1-1}\delta_1$$

is true for $0\leq i+i_1+\cdots+i_{l-1}<n$ and $k=i+i_1+\cdots+i_{l-1}$ then

$$C_{l+1}\rightarrow\gamma_{l+1}H_{n-k-i_l}(\alpha)D_{l+1}\varepsilon_{l+1} \text{ and}$$

$$D_{l+1}\overset{*}{\Rightarrow}T_{i_l}(H_{n-k}(\alpha))D_l\delta_{l+1}$$

do not exist together for i_l, $0\leq i+i_1+\cdots+i_{l-1}+i_l<n$. Furthermore (iv′) is extended to (iv″).

(iv″) If $A\rightarrow\alpha\beta$ and $B\rightarrow\beta$ exist in P and $Q_T(A)\cap Q_T(B)\neq\phi$ then

$$C_1\rightarrow\gamma_1 H_{n-i}(\alpha)D_1\varepsilon_1 \text{ and } D_1\overset{*}{\Rightarrow}T_i(\alpha)B\delta_1$$

do not exist together for i such that $Q_T(A)\cap Q_T(Ti(\alpha)|B)\neq\phi$. And if the latter is true for $0\leq i<n$ then

$$C_2\rightarrow\gamma_2 H_{n-i-i_1}(\alpha)D_2\varepsilon_2 \text{ and } D_2\overset{*}{\Rightarrow}T_{i_1}(H_{n-i}(\alpha))D_1\delta_2$$

do not exist together for i_1, $0\leq i+i_1<n$. By repeating this process for l times, if

$$D_l\overset{*}{\Rightarrow}T_{i_{l-1}}(H_{n-k+i_{l-1}}(\alpha))D_{l-1}\delta_l$$

is true for $0 \leq i+i_1+\cdots+i_{l-1}<n$, $k=i+i_1+\cdots+i_{l-1}$, then

$C_{l+1} \rightarrow \gamma_{l+1} H_{n-k-i_l}(\alpha) D_{l+1} \varepsilon_{l+1}$ and

$D_{l+1} \overset{*}{\Rightarrow} T_{i_l}(H_{n-k}(\alpha)) D_l \delta_{l+1}$ do not exist together.

In a right precedence grammar $G=(V_N, V_T, P, S)$ with conditions (i), (ii), (iii) and (iv″), and including the production rules $A \rightarrow \alpha\beta$ and $B \rightarrow \beta$ in P, the string $\alpha\beta$ is not necessarily a handle, even if the stack configuration becomes ... $\alpha\beta$ on a step of parsing of a sentence in $L(G)$, where $T_1(\beta) \gtrdot T$ and T is an input terminal symbol at this step.

In this case, if there is a precedence relation between B and T but not between A and T, then β may be a handle and $B \rightarrow \beta$ may be applied. This case does not appear under the condition (iv) because the production rule $D \rightarrow Ti(\alpha)B\delta$ is essentially forbidden under the condition (iv) in case of $i=n$, if both $A \rightarrow \alpha\beta$ and $B \rightarrow \beta$ exist in P. But if $Q_T(A) \cap Q_T(Ti(\alpha)|B)=\phi$ and $T \in Q_T(Ti(\alpha)|B) \cap Q_T(B)$ are satisfied then we do not need to remove $D \rightarrow Ti(\alpha)B\delta$ under the condition (iv″).

3.2 *On the Rule (ii)*

Under the condition (ii), it was unconditionally forbidden that $A \rightarrow \alpha$ and $B \rightarrow \alpha (A \neq B)$ exist together in P. But we may loosen a little the restriction of (ii).

(ii′) If $Q_T(A) \cap Q_T(B) \neq \phi$ $(A \neq B)$ is true then

$A \rightarrow \alpha$ and $B \rightarrow \alpha$ do not exist together in P.

3.3 *A Parsing Method for Extended Right Precedence Grammars*

The grammars with conditions (i), (ii′), (iii) and (iv″) are called ‘extended right precedence grammars’. In a parsing method for simple right precedence grammars, we have taken the following procedure: at finding out the rightmost symbol of a handle using the right precedence table, we select the production having the longest right hand side which is equivalent to the top part of the stack and then apply it.

In this extended method, if both right hand sides of the production rules, $A \rightarrow \alpha\beta$ and $B \rightarrow \beta$ are equal to the string of the top part of the stack, and there is no production rule.

$C \rightarrow \gamma\alpha\beta$ $(|\gamma| \neq 0)$, where $\gamma\alpha\beta$ equals to the top part of the stack, then $A \rightarrow \alpha\beta$ is adapted and $\alpha\beta$ is the handle, when $A \gtrdot T$ or $A \lesseqdot T$ with $T_1(\beta) \gtrdot T$ for input symbol T.

In this case, even if $B \gtrdot T$ or $B \lesseqdot T$ exist,

$Q_T(A) \cap Q_T(B) \neq \phi$ is satisfied.

Therefore β is not the handle due to (iv′) if

$Q_T(A) \cap Q_T(Ti(\alpha)|B) \neq \phi$.

And if there is a precedence relation between B and T but not between A and T, then β is the handle, and then $B \rightarrow \beta$ is adapted.

References

[1] K. Inoue, 'Right Precedence Grammars', *IPSJ* 11, No. 8, 449 (Aug., 1971).
[2] J. Feldman and D. Gries, 'Translator Writing Systems', *Comm. ACM* 11, No. 2, 77 (Feb., 1968).