COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE

15,4 (1974)

ОБ ЭКВИВАЛЕНТНОСТИ τ_m -ГРАММАТИК И $Sb(n)$-ГРАММАТИК

Виктор АЛАДЬЕВ, Таллин

Резюме: В настоящей заметке доказывается строгая эквивалентность τ_m-грамматик и $Sb(n)$-грамматик.

Ключевые слова: Однородная структура; τ_m-грамматика; язык, генерируемый τ_m-грамматикой; $Sb(n)$-грамматика; язык, генерируемый $Sb(n)$-грамматикой; строгая эквивалентность; фигурная унарная алгебра.

AMS : 94A20, 68A30, 92A05 Ref. Ž.: 8.764.11

Основным объектом применения теории формальных грамматик являются не произвольные грамматики, а грамматики некоторых специальных типов, наиболее важных как в теоретическом, так и в практическом аспекте. Выделение этих типов производится по виду правил вывода. В последние годы большое внимание уделяется интенсивному изучению параллельных грамматик (см. библиографию в работах [1 - 4]). Такие грамматики обладают тем свойством, что на каждом шаге в выводе перерабатывается каждый символ слова, с которым оперирует грамматика, и представляют большой интерес для вычислительной техники, теории алгоритмов и программирования, а также биоматематики. Настоящая заметка является непосредственным продолжением работ [3 - 5] и выясняет степень общности понятия τ_m -грамматик, введенных автов работе [3]. Грамматики такого типа описывают на формальном уровне функционирование однородных структур, на которых реализуются вычислительные и логические устройства, а сами структуры весьма удобны в качестве технической базы для организации параллельной обработки информации [1].

В недавней своей работе [5] Е. Щербаков с целью более адекватного описания биологических процессов развития ввел в рассмотрение специальный класс алгебраических систем – фигурные унарные алгебры. Более того, он предположил, что понятие фигурной унарной алгебры является более общим, чем понятие однородной структуры. В настоящей заметке мы покажем, что даже более общие грамматики, чем грамматики, определяемые на основе фигурных унарных алгебр, эквивалентны τ_m - грамматикам. Все определения, понятия и обозначения, кроме вновь вводимых, полностью соответствуют работе [4].

Две грамматики будем называть строго эквивалентными, если они не только порождают один и тот же язык, но и порождают его идентичным образом.

Определние τ_m -грамматик можно найти в любой из работ [1 - 4].

$S\mathscr{b}(m)$ -грамматики (используя обозначения работы [5]) введем следующим образом. Имеется связанное множество GS идентичных конечных автоматов Мура, помещенных в точки пространства Z^1 [4]. Связь автоматов, не нарушая общности, определяется индексом соседства единичного A_i -автомата из множества $\vartheta = \{0, 1, \ldots, m-1\}$. Индекс соседства одинаков для всех автоматов и определяет шаблон соседства (набор соседних A_i -автомату автоматов) ШС =

$= \{i+0, i+1, \dots, i+m-1\}$, т.е. соседними A_i -автомату служат A_j -автоматы $(j = i + k;\ k = \overline{0, m-1})$. Каждый A_i -автомат может находиться в любой дискретный момент времени $T \geq 0$ только в одном состоянии из множества $S^p = \{0, 1, \dots, p-1\}$. Отображение $\theta : Z^1 \longrightarrow S^p$ будем называть словом. Слово, имеющее только конечное множество символов из $S^p \setminus \{0\}$ будем называть конечным и множество всех таких слов обозначать через $\overline{C}_p$. Мы будем рассматривать только слова из множества $\overline{C}_p$. Слово в момент $T = 0$ будем обозначать через c_0. Для порождения новых слов из слова c_0 вводятся соотношения вида

$$
(1)\quad \left.\begin{array}{l}
X_1 X_2 X_3 \dots X_m \longrightarrow Y_1 Y_2 Y_3 \dots Y_t \\
\underbrace{0\ 0\ 0 \dots 0}_{m} \longrightarrow \underbrace{0\ 0\ 0 \dots 0}_{t} \\
X_i, Y_j \in S^p\ (i = \overline{1, m};\ j = \overline{1, t};\ 1 \leq t \leq m),
\end{array}\right\} Sl
$$

где X_i есть состояния автоматов из множества GS, расположенных согласно шаблону соседства. В дальнейшем под автоматом A мы будем понимать не только его местоположение в GS, но и его состояние. Одновременное применение соотношений (1) определяет функцию Sl, т.е. для данного множества S^p, индекса соседства ϑ и слов $c_1, c_2 \in \overline{C}_p$ функция Sl должна определять преобразование слов $Sl: c_1 \longrightarrow c_2$. Но так как соотношения (1) применяются одновременно и правые их части длиной ≥ 1, то возможна ситуация, когда в какой-то момент времени $T > 0$ автомат A из множества GS должен будет иметь несколько внутренних состояний, что недопустимо. Поэтому соотношения (1) мы должны дополнить однозначными функциями выбора состояния

(2) $$\varphi(Z_1, Z_2, Z_3, \dots, Z_q)$$

$$(Z_i \in S^n;\ i = \overline{1, q};\ 1 \le q \le n)$$

со значениями во множестве S^n. Например, при $n = 2$, $m = 3$ одновременное применение соотношений

$000 \to 000$ $\quad$ $011 \to 101$

$001 \to 1$ $\quad$ $101 \to 10$

$010 \to 01$ $\quad$ $110 \to 01$

$100 \to 11$ $\quad$ $111 \to 010$

к слову $c_0 \in \overline{C}_2$ следующего вида

... 0000101110010000 ...

приводит к следующему результату

```
     ... 0 1 2 3 4 5 6 7 8 9 10 11 12 13 ...
     ... 0 0 0 1 0 1 1 1 0 0 1  0  0  0 ...
     ... 0 0 0
           1 - -
             0 1 -
               1 0 -
                 1 0 1
                   0 1 0
                     0 1 -
                       1 1 -
                         1 - -
                           0 1 -
                             1 1 -
                               0 0 0
...0 0 0 1 0 1 1 1 0 0 1 1 0 1 0 0 0 ...
```

т.е. в автоматах с номерами 1, 4, 6, 7, 11 из множества GS наблюдается неоднозначность при определении внутренних состояний, которая исчезает при определении функции выбора состояния

$\varphi(0,1) = 1$, $\qquad$ $\varphi(1,1,0) = 0$

$\varphi(1,0) = 0$ $\qquad$ $\varphi(0,1,1) = 1$.

При сделанных предположениях из начального слова $c_0 \in \overline{C}_2$ с помощью функции Sb_φ, определенной соотношениями (1)-(2), генерируется последовательность слов

$$\langle c_0 \rangle = c_0, c_1, c_2, \dots, c_i, \dots \qquad (c_i = c_0 (Sb_\varphi)^i;\ i = 1, 2, \dots).$$

Для случая наших грамматик под строгой эквивалентностью будет пониматься генерация грамматиками идентичных последовательностей слов из одной и той же аксиомы c_0. На этом же примере становится довольно прозрачным сам принцип одновременного применения к любому перерабатываемому слову соотношений (1).

Тогда $Sb(m)$ <u>-грамматикой</u> будем называть упорядоченную четверку $Sb(m) = (S^n, c_0, Sb, \varphi)$, где:

S^n - терминальный словарь;

$c_0 \in \overline{C}_n$ - начальное слово или аксиома;

Sb - параллельные преобразования;

φ - функция выбора состояния.

Функции Sb и φ определяют правила вывода Sb_φ в $Sb(m)$ -грамматике. Множество слов $\langle c_0 \rangle$, генерируемое такой грамматикой, будем называть $Sb(m)$ <u>-языком</u>. От общеизвестных формальных грамматик $Sb(m)$ -грамматики, также как и τ_m -грамматики, отличаются отсутствием нетерминального словаря и одновременным применением соотношений (1). Если для нашего определения $Sb(m)$ -грамматик положить в соотношениях (1) $m = t = 3$, а для функции φ положить

$$\varphi(0, 0) = 0 \qquad \varphi(X, 0) = X$$

$$\varphi(0, X) = X \qquad \varphi(X, Y) = * \notin S^p$$
$$Y, X \in S^p \setminus \{0\},$$

то мы приходим к случаю, рассмотренному в работе [5]. Очевидно, что любая τ_m -грамматика строго эквивалентна $Sb(m)$ -грамматике, у которой соотношения Sb имеют следующий вид

$$X_1 X_2 X_3 \dots X_m \longrightarrow X_1' \underbrace{0 \dots\dots 0}_{m-1}$$
$$X_1', X_i \in S^p \qquad (i = \overline{1, m}),$$

а функция выбора состояния имеет вид

$$\varphi(Z_1, Z_2, Z_3, \dots, Z_m) = \sum_{i=1}^{m} Z_i \pmod p$$
$$Z_i \in S^p \qquad (i = \overline{1, m}).$$

Покажем теперь, что и в $Sb(m)$ -грамматике существует строго эквивалентная ей τ_m -грамматика. Точнее, имеет место следующая

<u>Теорема</u>. Для каждой $Sb(m)$ -грамматики существует строго эквивалентная ей τ_m -грамматика $(m = 2m - 1)$. Если в соотношениях (1) величины m, $t \geq 1$ и произвольны, то величина $m = \max_i \{m_i\} + \max_i \{t_i\} - 1$.

<u>Доказательство</u>. Пусть имеется произвольная $Sb(m)$ -грамматика, порождающая из некоторй аксиомы $c_0 \in \overline{C}_p$ $Sb(m)$ -язык $G = \langle c_0 \rangle = c_0 \rightarrow c_1 \rightarrow c_2 \rightarrow \dots\dots \rightarrow c_i \rightarrow \dots$ $(c_j \in \overline{C}_p;\ j = 0, 1, \dots, i, \dots)$. Из определения $Sb(m)$ -грамматик следует, что они подобно τ_m -грамматикам, являются

алгоритмами, а не исчислениями. Следовательно, из любого предшественника $c_i \in \overline{C}_p$ мы всегда получаем только одного последователя $c_{i+1} \in \overline{C}_p$ с помощью правил вывода Sb_φ $Sb(m)$-грамматики. Докажем теперь, что существует τ_{2m-1}-грамматика с аксиомой $c_0 \in \overline{C}_p$, генерирующая ту же последовательность слов $\langle c_0 \rangle$, что и $Sb(m)$-грамматика. Вообще говоря, такое ограничение как рассмотрение только конечных слов в доказательстве теоремы никак не используется и может быть снято, если рассматривать моделирование $Sb(m)$-систем однородными структурами. Более того, снятие с соотношений (1) ограничения $\underbrace{000\ldots0}_{m} \longrightarrow$ $\longrightarrow \underbrace{000\ldots0}_{t}$ также не изменяет справедливости доказательства нашей теоремы. Таким образом, $Sb(m)$-грамматика любое слово $c_d \in \overline{C}_p$ переводит в слово $c_{d+1} \in \overline{C}_p$. Рассмотрим произвольный отрезок ФА длиной $2m-1$ слова c_d

$$c_d = \ldots\ldots x^d_{i+1}\ x^d_{i+2} \ldots\ldots \boxed{x^d_{i+m}}\ x^d_{i+(m+1)} \ldots\ldots x^d_{i+2m-1} \ldots\ldots$$

$$Sb \begin{cases} \ldots\ldots y^1_{i+1}\ y^1_{i+2} \ldots\ldots y^1_{i+m} \\ \qquad\qquad y^2_{i+2} \ldots\ldots y^2_{i+m}\ y^2_{i+(m+1)} \\ - \ - \ - \ - \ - \ - \ - \ - \ - \ - \\ - \ - \ - \ - \ - \ - \ - \ - \ - \ - \\ \qquad\qquad\qquad\quad y^m_{i+m} \ldots\ldots\ldots\ldots y^m_{i+2m-1} \ldots\ldots \end{cases}$$

$$c_{d+1} = \ldots\ldots\ldots\ \varphi(y^1_{i+m}, y^2_{i+m}, \ldots, y^m_{i+m}) \ldots\ldots$$

В результате одновременного применения соотношений (1) к слову c_d происходит наложение внутренних состояний автоматов из множества GS $\overline{\psi^{l}_{i+k} \ (l = 1, m}$; $\overline{k = i, i + m - 1})$. Выбираем в отрезке ФА слова c_d автомат X^{d}_{i+m} в качестве центрального. Под действием одновременных соотношений (1) и функции выбора состояния φ состояние центрального автомата X^{d}_{i+m} в момент $T = d+1$ (в слове c_{d+1}) будет определяться значением $\varphi(\psi^{1}_{i+m}, \psi^{2}_{i+m}, \ldots, \psi^{m}_{i+m}) \in S^{p}$ при $\psi^{k}_{i+m} \in S^{p} (k = \overline{1,m})$. Более того, это значение уже не зависит от того, какие применяются соотношения Sb к слову c_d справа от автомата X^{d}_{i+m} и слева от автомата X^{d}_{i+1}. Из вышесказанного возникает следующий способ построения τ_m-грамматики, генерирующей тот же язык $G = \langle c_0 \rangle = c_0 \longrightarrow c_1 \longrightarrow c_2 \longrightarrow \ldots$ $\ldots \longrightarrow c_\theta \longrightarrow \ldots \ (\theta = \theta + 1)$. Прежде всего необходимо положить $m = 2m - 1$, ибо это есть минимальный размер шаблона соседства τ_m-грамматики, позволяющий полностью получить информацию для определения состояния любого автомата X^{d+1}_{i+m} в слове c_{d+1} при его выводе. Правила же вывода $L^{\vartheta}_{(2m-1)}$ [4] в такой τ_{2m-1}-грамматике определяем следующим образом. К произвольным наборам

$$X^{d}_{i+1} \ X^{d}_{i+2} \ldots\ldots X^{d}_{i+2m-1}, \quad (X^{d}_{i+j} \in S^{p};\ j = \overline{1, 2m-1})$$

применяем одновременно соотношения (1) как это мы уже делали для случая слова c_d. Очевидно, что можно применить не более m различных соотношений из (1). В результате такой операции мы получаем не более p различных символов $y^{k}_{i+m} \in S^{p}$ $(k = \overline{1, \omega \leq m})$ для определения значения автомата X^{d+1}_{i+m} в слове c_{d+1}, т.е. полагаем значение функции $L^{\vartheta}_{(2m-1)}$ равным

$$L^{\vartheta}_{(2m-1)}(X^{d}_{i+m}, X^{d}_{i+m+1}, \dots, X^{d}_{i+2m-1}) = g(y^{1}_{i+m}, y^{2}_{i+m}, \dots, y^{k}_{i+m}).$$

Таким образом, мы полностью задаем τ_{2m-1} -грамматику. Используя вышеизложенное уже нетрудно показать, что такая τ_{2m-1} -грамматика генерирует в точности язык

$$G = \langle c_0 \rangle = c_0 \longrightarrow c_1 \longrightarrow c_2 \longrightarrow \dots \longrightarrow c_i \longrightarrow \dots .$$

Этим первая часть теоремы доказана. Пусть теперь при определении $Sb(m)$ -грамматик в соотношениях (1) величины $m, t \geq 1$ и являются произвольными. В этом случае идея доказательства полностью сохраняется с очевидными изменениями: если в первом случае центральный автомат отрезка ФА имел в эквивалентной τ_m -грамматике индекс соседства $\vartheta = \{-(m-1), \dots, -1, 0, 1, \dots, (m-1)\}$, то в этом случае индекс соседства его имеет вид

$$\vartheta = \{-(t-1), \dots, -1, 0, 1, \dots, (m-1)\},$$

где $t = \max_i (t_i)$ и $m = \max_i (m_i)$. Остальные рассуждения в сущности не изменяются. Этим теорема полностью доказана.

Из нашей теоремы следует, что понятие τ_m -грамматик (однородных структур) является довольно общим и фигурные унарные алгебры не приводят к обобщению этого понятия.

А так как τ_m -грамматики, на наш взгляд, более удобны с точки зрения изучения, использования и интерпретации, то имеет смысл уделить внимание изучению именно этого типа грамматик.

Литература

[1] V. ALADYEV: Survey of research in the theory of homogeneous structures and their applications, Mathemat.Biosci.,Amer.Elsevier Publ.Co.,Inc.1974.

[2] A. SALOMAA: Formal languages, Academic Press,1973.

[3] В. АЛАДЬЕВ: τ_m -грамматики и порождаемые ими языки, Изв.АН ЭССР,Биология,23(1974),67-87.

[4] В. АЛАДЬЕВ: Операции над языками, генерируемыми τ_m-грамматиками, Comment.Math.Univ.Carolinae 15(1974),211-220.

[5] Е. ЩЕРБАКОВ: О фигурных операциях параллельной подстановки и порождаемых ими унарных алгебрах, Вычислительная техника и вопросы кибернетики, Изд-во Ленинградского университета,вып.10 (1974),90-99.

Эстонский филиал ВГПТИ ЦСУ СССР
Таллин 200001
Маакри 15
СССР

(Oblatum 23.8.1974)