Generative Power of Context-Conditional Grammars

R.O. Oladele* and B. Yusuf

Department of Computer Science, University of Ilorin, PMB 1515, Ilorin, Nigeria.

E-mail: roladele@unilorin.edu.ng*
yb4real@gmail.com
letua034@yahoo.com

ABSTRACT

This paper studies the generative power of context-conditional grammars. It is proved that every recursively enumerable language can be generated by a context-conditional grammar of degree (1, 1) with no more than 5 conditional productions and 7 non-terminals. The result is based on the idea of using the so called Geffert normal forms for type-0 grammars.

(Keywords: formal languages, generative power, context-conditional grammars)

INTRODUCTION

Context-free grammars are the most investigated type of Chomsky hierarchy. They have good mathematical properties and are widely used in many practical applications. They were first used to study human languages. More recent applications include the structure of mark-up languages such as HTML and XML. Another important application of context-free grammars occurs in programming languages specification and compilers design. However, context-free grammars cannot cover all aspects of natural and programming languages.

Consequently, some more powerful regulating mechanisms that generate convenient proper sub-class of the class of context-sensitive language and that exploit the advantage of the simple form of context-free productions are of interest. It is therefore natural to reduce this regulating mechanism without any decrease of the generative power. This paper demonstrates this reduction in terms of context-conditional grammars.

It was proved in [2] that the class of recursively enumerable languages is generated by a context-conditional grammar of degree (2, 1) with seven conditional productions and eight non-terminals. This paper investigates the generative power of context-conditional grammars of degree (1, 1). We improve on the degree and both the number of non-terminals and the number of conditional productions. Specifically it is proved that the class of recursively enumerable languages is generated by a context-conditional grammar of degree (1, 1) with five conditional productions and seven non-terminals.

PRELIMINARIES

It is assumed in this paper that the reader is familiar with the theory of formal language (see [4, 5]). For an alphabet V, V^* denotes the free monoid generated by V. The unit of V^* is denoted by ε. Set $V^+ = V^* - \{\varepsilon\}$. For $w \in V^*, |w|$ represents the length of w. Set $sub(w) = \{ u : u \text{ is a subword of } w\}$. For a finite subset $W \subseteq V^*$, $\max(W)$ is the minimal non-negative integer n, such that $|x| \le n, \forall x \in W$.

In [1], it was proved that every recursively enumerable language is generated by a grammar $G = (\{S, S', A, B\}, T, P \cup \{ABBBA \to \varepsilon\}, S')$, where P consists of context-free productions of the forms:

$$S' \to uS'a, where\ u \in \{AB, ABB\}^*, a \in T$$

$$S' \to S$$

$$S \to uSv, where\ u \in \{AB, ABB\}^*, \quad v \in \{BBA, BA\}^*,$$

$$S \to uv, where\ u \in \{AB, ABB\}^*, \quad v \in \{BBA, BA\}^*.$$

With respect to the rules above, the derivation of a word $w \in T^*$ can be divided into three stages: in the first stage $S' \overset{*}{\Rightarrow} w'S'w \Rightarrow w'Sw$ where $w' \in \{AB, ABB\}^*$ and $w \in T^*$ is obtained by using productions of the form $S' \rightarrow uS'a$ and $S' \rightarrow S$, where $u \in \{AB, ABB\}^*$ and $a \in T$, in the second stage $w'Sw \overset{*}{\Rightarrow} w_1w_2w$, where $w_1 \in \{AB, ABB\}^*$ and $w_2 \in \{BBA, BA\}^*$ is obtained by using productions of the form $S \rightarrow uSv$ and $S \rightarrow uv$, where $u \in \{AB, ABB\}^*$, and $v \in \{BBA, BA\}^*$, and in the third stage $w_1w_2w \overset{*}{\Rightarrow} w$ is obtained by only using $ABBBA \rightarrow \varepsilon$.

A context-conditional grammar, G is a quadruple $G = (N, T, P, S)$ where N is a non-terminal alphabet, T is a terminal alphabet, $N \cap T = \emptyset, S \in N$ is the start symbol, and P is a finite set of productions of the form $(X \rightarrow \alpha, Per, For)$ where $X \in N$, $\alpha \in (N \cup T)^*$, and $Per, For \subseteq (N \cup T)^+$ are finite sets. If $Per \cup For$ is a non-empty set, then the production is said to be conditional. G has degree (i, j) if for all productions $(X \rightarrow \alpha, Per, For) \in P, \max(Per) \leq i \text{ and } \max(For) \leq j$.

For $x \in (N \cup T)^+ \text{ and } y \in (N \cup T)^*, x$ directly derives y according to the production $(X \rightarrow \alpha, Per, For) \in P$, denoted as $x \mapsto y, \text{if } x = x_1Xx_2,\ y = x_1\alpha x_2$ for some $x_1, x_2 \in (N \cup T)^*, Per \subseteq sub(x) \text{ and } For \cap sub(x) = \emptyset$.

The language generated by a context-conditional grammar, G, is defined as $L(G) = \{w \in T^* : S \overset{*}{\Rightarrow} w\}$.

MAIN RESULT

The main result of this paper is presented in this section.

Theorem 1: Every recursively enumerable language is generated by a context conditional grammar of degree (1, 1) with no more than 5 conditional productions and 7 non terminals.

Proof: Let L be a recursively enumerable language. Then, there is a grammar:

$G_1 = (\{S, S' A, B\}, T, P \cup \{ABBBA \rightarrow \varepsilon\}, S')$

in the Geffert normal form such that $L = L(G1)$. Construct the grammar:

$G_2 = (\{S, S', A, B, A', B', B"\}, T, P' \cup P", S')$,

where

$P' = \{(X \rightarrow \alpha, \emptyset, \emptyset) : X \rightarrow \alpha \in P\}$
and $P"$ contains the following five conditional productions:

1. $(A \rightarrow A', \emptyset, \{A'\})$

2. $(B \rightarrow B', \emptyset, \{B', B"\})$

3. $(A' \rightarrow \varepsilon, \{B"\}, \emptyset)$

4. $(B' \rightarrow B", \emptyset, \{B"\})$

5. $(B" \rightarrow \varepsilon, \emptyset, \{B'\})$

To prove $L(G_1) \subseteq L(G_2)$, consider a derivation $S' \Rightarrow^* uABBBAvw \Rightarrow uvw$ in G_1 by productions from P and the only one production $ABBBA \rightarrow \varepsilon$ where $v, w \in \{A, B\}^*$ and $w \in T^*$ Then, $S' \Rightarrow^* uABBBAvw$ in G_2 by productions from P'. Further, by productions from P'' we obtain:

$$\begin{aligned} \mathrm{uABBBAvw} &\overset{(1)}{\Rightarrow} \mathrm{uA'BBBAvw} \\ &\overset{(2)}{\Rightarrow} \mathrm{uA'B'BBAvw} \\ &\overset{(4)}{\Rightarrow} \mathrm{uA'B"BBAvw} \\ &\overset{(3)}{\Rightarrow} \mathrm{uB''BBAvw} \\ &\overset{(5)}{\Rightarrow} \mathrm{uBBAvw} \\ &\overset{(2)}{\Rightarrow} \mathrm{uB'BAvw} \end{aligned}$$

$\overset{(4)}{\Rightarrow}$ uB″BAvw

$\overset{(5)}{\Rightarrow}$ uBAvw

$\overset{(2)}{\Rightarrow}$ uB′Avw

$\overset{(4)}{\Rightarrow}$ uB"Avw

$\overset{(1)}{\Rightarrow}$ uB"A′vw

$\overset{(3)}{\Rightarrow}$ uB"vw

$\overset{(5)}{\Rightarrow}$ uvw

The inclusion follows by induction.

To prove $L(G_1) \subseteq L(G_2)$, consider a terminal derivation. Notice that to eliminate a nonterminal, there is B'' in the derivation. From production 4 and the observation that there is no more than one A', B' in the derivation (see productions 1, 2), there cannot be a terminal between any two non-terminals.

Therefore, the derivation is of the form $S \Rightarrow * \; w_1 w_2 w_3$ in G_1 by productions from P', where

$w_1 \in \{AB, ABB\}^*$,
$w_2 \in \{BBA, BA\}^*$, $w_3 \in T^*$,
and $w_3 \in T^*$,
and $w_1 w_2 w_3 \Rightarrow^* w_3$.

Note that before S is eliminated, there is no occurrence of the substring $ABBBA$ in the derivation. Then, $S \Rightarrow^* w_1 w_2 w_3$ in G_1 by productions from P. We prove that $w_1 w_2 w_3 \Rightarrow^* w_3$ in G_1.

For $w_1 w_2 = \varepsilon$, the proof is done.

If $w_1 w_2 \neq \varepsilon$, there is B in $w_1 w_2$: otherwise, B'' cannot be obtained and no nonterminal can be eliminated. To obtain B'' production 4 is applied. Therefore, $w_1 w_2 = uABBBAv$ where $u \in \{AB, ABB\}^*$ and $v \in \{BBA, BA\}^*$ otherwise, the conditions of production 4 are not met.

Thus, at the beginning, only productions 1 and 2 are applicable.

$$
\begin{aligned}
\mathrm{uABBBAvw_3} &\overset{2}{\Rightarrow} \mathrm{uA'B'BBAvw_3} \\
&\Rightarrow \mathrm{uA'B''BBAv} \\
&\overset{3}{\Rightarrow} \mathrm{uB'BAvw_3} \\
&\Rightarrow \mathrm{uB''BAvw_3} \\
&\overset{2}{\Rightarrow} \mathrm{uB'Av} \\
&\Rightarrow \mathrm{uB''Avw_3} \\
&\overset{2}{\Rightarrow} \mathrm{uB''vw_3} \Rightarrow \mathrm{uvw_3}
\end{aligned}
$$

Thus, if $S \Rightarrow^* w_1 w_2 w_3$ in G_2, where $w_1 \in \{AB, ABB\}^*$, $w_2 \in \{BBA, BA\}^*$, $w_3 \in T^*$

Then, $S \Rightarrow^* w_1 w_2 w_3 \Rightarrow^* w_3$ in G_1.

REFERENCES

1. Geffert, V. 1988. “Context-Free-Like Forms for the Phrase-Structure Grammars”. In: M. Chytil, L. Janiga, and V. Koubek (eds.). *MFCS. Volume 324 of Lecture Notes in Computer Science,* Springer: Berlin, Germany. 324:309–317.
2. Masopust, T. and A. Meduna. 2007. “Descriptional Complexity of Grammars Regulated by Context Conditions”. *Proceedings on Language and Automata Theory (LATA)*. 403-412.
3. Okubo, F. 2009. “A Note on the Descriptional Complexity of Semi-Conditional Grammars”. *Information Processing Letters*. 110:36–40.
4. Paun, G. 1985. “A Variant of Random Context Grammars: Semi-Conditional Grammars”. *Theoret. Comput. Sci.* 41:1–17.
5. Salomaa, A. 1973. *Formal Languages*. Academic Press: New York, NY.
6. Turaev, S.G.M. 2012. “Descriptional Complexity of Lindenmayer Systems”. *App. Math. and Comp. Intel.*. 1:12–23.

SUGGESTED CITATION

Oladele, R.O. and B. Yusuf. 2017. "Generative Power of Context-Conditional Grammars". *Pacific Journal of Science and Technology*. 18(2):128-131.

Pacific Journal of Science and Technology