SCATTERED CONTEXT GRAMMARS CAN GENERATE THE POWERS OF 2

Tomáš Masopust
Doctoral Degree Programme (3), FIT BUT
E-mail: masopust@fit.vutbr.cz

Supervised by: Alexander Meduna
E-mail: meduna@fit.vutbr.cz

ABSTRACT

In this paper, a scattered context grammar generating all powers of 2 is given.

1 INTRODUCTION

Scattered context grammars were introduced by Greibach and Hopcroft in 1969 (see [1]). Since their introduction, many properties concerning them have been proved. However, there still remain open questions. To mention one of the most important of them, consider the Chomsky hierarchy, i.e. families of regular, context-free, context sensitive, and recursively enumerable languages. It is not hard to prove that all scattered context languages (defined in [1], i.e. without epsilon-productions) are context sensitive. The proof is based on the so-called workspace theorem (see [3, Theorem III.10.1]). Thus, if we denote by SC and CS the families of all scattered context and context sensitive languages, respectively, it is well-known that

$$SC \subseteq CS.$$

One of the most famous open problem concerning scattered context grammars is whether there is a language that is context sensitive and is not scattered context. In other words, is it true that

$$SC \subsetneq CS?$$

A few languages seemed to be able to prove this inequality. One of them was the language

$$\{a^{2^n} : n \geq 0\}.$$

However, in this paper, we show that this language does not prove the inequality.

2 DEFINITIONS

We assume that the reader is familiar with the theory of formal languages (see [2]). An alphabet is an arbitrary finite set. For an alphabet V, V^* represents the free monoid generated by V. The unit of V^* is denoted by ε. Set $V^+ = V^* - \{\varepsilon\}$. Let CS denote the family of all context sensitive languages.

A *scattered context grammar* (*SCG*, for short) is a quadruple

$$G = (V, T, P, S),$$

where

- V is a total alphabet,
- $T \subseteq V$ is a terminal alphabet,
- $S \in V - T$ is the start symbol, and
- P is a finite set of productions of the form

$$(A_1, \ldots, A_n) \to (x_1, \ldots, x_n),$$

  for some $n \geq 1$, where $A_i \in V - T$ and $x_i \in V^+$, for $1 \leq i \leq n$.

If $(A_1, \ldots, A_n) \to (x_1, \ldots, x_n) \in P$, $u = u_1A_1u_2 \ldots u_nA_nu_{n+1}$, and $v = u_1x_1u_2 \ldots u_nx_nu_{n+1}$, where $u_i \in V^*$, for $1 \leq i \leq n$, then

$$u \Rightarrow v \; [(A_1, \ldots, A_n) \to (x_1, \ldots, x_n)]$$

in G or, simply, $u \Rightarrow v$. Let $\Rightarrow^+$ and $\Rightarrow^*$ denote the transitive and the transitive and reflexive closures of $\Rightarrow$, respectively. The *language* of G is denoted by $\mathscr{L}(G)$ and defined as

$$\mathscr{L}(G) = \{x \in T^* : S \Rightarrow^* x\}.$$

A language, L, is a scattered context language if there is a scattered context grammar, G, such that $L = \mathscr{L}(G)$. Let *SC* denote the family of all scattered context languages.

3 MAIN RESULT

In this section, a construction of a scattered context grammar generating the set of all powers of 2 is given. The construction can easily be modified to generate the set of all powers of k, for some fixed $k \geq 2$.

Theorem 1. *There is a SCG, G, such that* $\mathscr{L}(G) = \{a^{2^n} : n \geq 0\}$.

Proof. Define $G = (\{S, A, A', X, Y, Z\}, \{a\}, P, S)$, where P contains

1. $(S) \to (a)$
2. $(S) \to (a^2)$
3. $(S) \to (A'AXY)$
4. $(A', A, X, Y) \to (a, A', X, A^2Y)$
5. $(A', X, Y) \to (a, A', AXY)$
6. $(A', X, Y) \to (Z, Z, a)$
7. $(Z, A, Z) \to (Z, a, Z)$
8. $(Z, Z) \to (a, a)$

Consider a string of the form

$$a^*A'A^nXA^mY,$$

where $m,n \geq 0$. By production 6,

$$a^*A'A^nXA^mY \Rightarrow a^*ZA^nZA^ma \quad [6].$$

Then, we can either finish the derivation by production 8 (if $n = 0$), or continue by production 7 as long as possible, followed by production 8;

$$a^*ZA^nZA^ma \Rightarrow^* a^*Za^nZA^ma \Rightarrow a^*aa^naA^ma \quad [7^*8].$$

It is easy to see that

$$a^*aa^naA^ma \in L \text{ if and only if } m = 0.$$

Therefore, to apply production 6, the sentential form must be of the form $a^*A'A^nXY$, for some $n \geq 1$ (from the previous and production 3).
If production 5 is applied to $a^*A'A^nXA^mY$,

$$a^*A'A^nXA^mY \Rightarrow a^*aA^nA'A'A^mAXY \quad [5],$$

then the derivation is blocked; indeed, A on the left of A' cannot be removed. Thus, if $m \geq 1$, the only possible derivation is to apply production 4 as long as there is A between A' and X.
Now, from the previous follows that if the first production applied to

$$a^{2^n-2}A'A^{2^n-1}XY$$

is production 4, then the derivation continues by production 4 as long as there is A between A' and X, followed by one application of production 5,

$$a^{2^n-2}A'A^{2^n-1}XY \Rightarrow^+ a^{2^n-2}a^{2^n-1}A'XA^{2^n-1}A^{2^n-1}Y \Rightarrow a^{2^{n+1}-2}A'A^{2^{n+1}-1}XY \quad [44^*5],$$

where $n \geq 1$.
Now, the derivation can successfully rewrite all A's to a, by productions 6, 7, and 8, i.e.

$$a^{2^n-2}A'A^{2^n-1}XY \Rightarrow^* a^{2^n-2}aa^{2^n-1}aa = a^{2^{n+1}} \quad [67^+8],$$

or continue by production 4. Then, it follows by induction.
Finally, we summarize all possible terminal derivations.

$$\begin{array}{lll}
S \Rightarrow & a & [1] \\
S \Rightarrow & a^2 & [2] \\
S \Rightarrow & A'AXY & [3] \\
\Rightarrow^* & a^2A'A^3XY & [44^*5] \\
\Rightarrow^* & a^6A'A^7XY & [44^*5] \\
\vdots & & \\
\Rightarrow^* & a^{2^{n-1}-2}A'A^{2^{n-1}-1}XY & [44^*5] \\
\Rightarrow^* & a^{2^n} & [67^+8]
\end{array}$$

Thus, any derivation generating a terminal string, w, is in one of the following three forms:

$$S \Rightarrow^* w\ [1] \quad \text{or} \quad S \Rightarrow^* w\ [2] \quad \text{or} \quad S \Rightarrow^* w\ [3(44^*5)^*67^+8].$$

□

4 OPEN PROBLEM

Although the previous language was proved to be scattered context, the question whether

$$SC = CS$$

or not remains open. However, another interesting nontrivial potential language that seems to be complicated enough for scattered context grammars to generate is the set of all prime numbers, i.e. the context sensitive language

$$\{a^p : p \text{ is a prime number}\}.$$

Whether this language is or is not scattered context is a question of the next research.

ACKNOWLEDGEMENTS

This work was supported by the Czech Grant Agency project No. 102/05/H050, FRVŠ grant No. FR762/2007/G1, and the Czech Ministry of Education under the Research Plan No. MSM 0021630528.

REFERENCES

[1] S. Greibach and J. Hopcroft. Scattered context grammars. *Journal of Computer and System Sciences*, 3:233–247, 1969.

[2] A. Meduna. *Automata and Languages: Theory and Applications*. Springer-Verlag, London, 2000.

[3] A. Salomaa. *Formal languages*. Academic Press, New York, 1973.