

AN OBSERVATION ON SCATTERED GRAMMARS

by

Andrzej Ehrenfeucht
Department of Computer Science
University of Colorado at Boulder
Boulder, Colorado

Grzegorz Rozenberg
Department of Mathematics
University of Antwerp, U.I.A.
Wilrijk, Belgium

#CU-CS-135-78

September, 1978

ALL Correspondence to second author.

- WORKING PAPER -

Scattered context-grammars were introduced in [1]. They constitute one of the most interesting examples of string rewriting systems that are "context-dependent". However they form one of the least understood types of grammars. Although the relationship of (some of the variations of) the class of scattered context grammars to some classes of languages was studied (see e.g. [2], [3]) the main question remains open: Do scattered context grammars generate all context sensitive languages? Clearly an answer to this question would shed light on the role that context plays in grammars.

In this work we show that recursively enumerable languages possess a very strong representation in the class of scattered context languages. As a matter of fact it is well known that they possess the same representation in the class of context sensitive languages!!! Thus a result of this form supports in some sense the conjecture that the class of scattered context languages equals the class of context sensitive languages. In any case it says that the class of scattered context languages is "close to" the class of context sensitive languages. Since we provide a direct proof of our result (an explicit construction is given) it gives some insight into how scattered grammars work.

We assume the reader to be familiar with Post normal systems (e.g. in the scope of [4]). If H is a Post normal system, ω is its axiom and $\alpha_1\$ \rightarrow \β_1 , $\alpha_2\$ \rightarrow \β_2 are its productions then we assume that $|\alpha_1| = |\alpha_2| < |\omega|$ and $|\beta_1|, |\beta_2| > 0$. The language of H consists of all the words over its alphabet for which there is no applicable production (it is all the words x generated in H with the property that whenever $\alpha\$ \rightarrow \β is a production in H then α is not a prefix of x). Clearly all recursively enumerable languages are generated by such systems. (In this note we consider two languages equal if they differ at most by the empty language.)

First let us recall the notion of a scattered context grammar.

A *scattered context grammar* is a 4-tuple $G = (V, \Sigma, P, S)$ where V is a finite nonempty alphabet, $\Sigma \subsetneq V$, $S \in V \setminus \Sigma$ and P is a finite nonempty set of productions each of which is of the form $(A_1, \dots, A_n) \rightarrow (w_1, \dots, w_n)$ where $n \geq 1$, $A_1, \dots, A_n \in V \setminus \Sigma$ and $w_1, \dots, w_n \in V^+$. For $\alpha = \alpha_0 A_1 \alpha_1 A_2 \dots A_n \alpha_n$ in Σ^+ , $\beta = \alpha_0 w_1 \alpha_1 \dots w_n \alpha_n$ and $(A_1, \dots, A_n) \rightarrow (w_1, \dots, w_n)$ in P we write $\alpha \xRightarrow{G} \beta$, and \xRightarrow{G}^* denotes the transitive and reflexive closure of the relation \xRightarrow{G} . The *language of G* is defined by $L(G) = \{\alpha \in \Sigma^* : S \xRightarrow{G}^* \alpha\}$; it is referred as a *scattered context language*.

Defining nontrivial languages by scattered context grammars is more difficult than by context sensitive grammars. As a matter of fact no examples of nontrivial scattered context grammars are known in the literature. For this reason we provide now an example of (what we believe is) a nontrivial scattered context language.

Example. Let $G = (V, \Sigma, P, S)$ be the scattered context grammar where $V = \{S, B_1, B_2, B_3, B_4, A, a\}$, $\Sigma = \{a\}$ and P consists of the following productions:

$$(B_1, A, B_2, B_3) \rightarrow (a, B_1, B_2, A^2 B_3),$$

$$(B_1, B_2, B_3) \rightarrow (a, B_1, B_2 B_3),$$

$$(B_1, B_2, B_3) \rightarrow (a, B_4, A),$$

$$(B_4, A) \rightarrow (B_4, a) \text{ and}$$

$$(B_4) \rightarrow (a).$$

One can prove that

$$L(G) = \{a^{2^{n+1} + n+1} : n \geq 1\}.$$

Here is our representation theorem for recursively enumerable languages. (For an alphabet V and its subalphabet Σ , $pres_{V, \Sigma}$ is the homomorphism on V^* that erases each letter from $V \setminus \Sigma$ and maps each letter from Σ into itself.)

Theorem. Let K be a recursively enumerable language over an alphabet Σ . There exists a scattered context language M such that $M \subseteq \{\phi\}^*K$ and $pres_{\Sigma \cup \{\phi\}, \Sigma} M = K$ where $\phi \notin \Sigma$.

Proof.

Let Δ be an alphabet such that $\#\Delta = \#\Sigma$, let f be a homomorphism from Σ onto Δ and let $K_\Delta = f(K)$. Clearly K_Δ is recursively enumerable.

Let H be a Post normal system generating K_Δ . Let ω be the axiom of H , V its total alphabet, $\bar{V} = \{\bar{a} : a \in V\}$, $\overline{\bar{V}} = \{\overline{\bar{a}} : a \in V\}$ and let $S, \phi, \bar{\phi}$ and $\hat{\phi}$ be new symbols.

Let $G = (Z, \Sigma, P, S)$ be a scattered context grammar such that $Z = V \cup \bar{V} \cup \{S, \phi, \bar{\phi}, \hat{\phi}\} \cup \Sigma$

and P consists of the following productions:

(1) $(S) \rightarrow (\bar{\phi} \gamma \bar{x})$ where $x \in V$ and $\omega = \gamma x$.

(2) For every production $a_1 \dots a_k \$ \rightarrow \$ b_1 \dots b_\ell$ in H and every x in V $(\bar{\phi}, a_1, \dots, a_k, \bar{x}) \rightarrow (\phi, \phi, \dots, \phi, \bar{\phi}, x b_1 \dots b_{\ell-1} \overline{\bar{b}}_\ell)$ is a production in P ,

(3) Let k be a positive integer such that for every production $\alpha \$ \rightarrow \$ \beta$ in H , $|\alpha| = k$ and let T be the set of all words over V of length k such that whenever $\alpha \$ \rightarrow \$ \beta$ is a production in H then $\alpha \notin T$. Then for every $d_1 \dots d_k$ in T , $d_1, \dots, d_k \in V$ and every x in V

$(\bar{\phi}, d_1, \dots, d_k, \bar{x}) \rightarrow (\phi, \phi, \dots, \phi, \hat{\phi} d_1 \dots d_k, \overline{\bar{x}})$

is a production in P .

(4) For every x, y in V

$(\hat{\phi}, y, \overline{\bar{x}}) \rightarrow (\hat{\phi}, f^{-1}(y), \overline{\bar{x}})$ and

$(\hat{\phi}, \overline{\bar{x}}) \rightarrow (\hat{\phi}, f^{-1}(x))$

are productions in P .

(5) $(\hat{\phi}) \rightarrow (\phi)$

is a production in P .

Now one can easily prove that indeed $L(G) \subseteq \{\phi\}^* K$ and $pres_{\Sigma \cup \{\phi\}, \Sigma}^M = K$.

The key observation is that

- (i) a production of type (2) can be applied (in a successful derivation) only to consecutive occurrences of a_1, \dots, a_k because otherwise one obtains a word where an occurrence of a letter from V is to the left of an occurrence of ϕ ; the situation that never leads to a terminal word.
- (ii) a word w is translated to a terminal word only if the word \tilde{w} corresponding to w in H is such that no production in H is applicable to \tilde{w} .

REFERENCES

- [1] S. Greibach and J. Hopcroft, Scattered context grammars, J. Comp. Syst. Sciences 3, 1969, 233-247.
- [2] O. Mayer, Some restrictive devices for context-free grammars, Information and Control, 20, 69-72, 1972.
- [3] D. Milgram and A. Rosenfeld, A note on scattered context grammars, Inf. Proc. Letters
- [4] M. Minsky, Computation: Finite and infinite machines, Prentice-Hall, Englewood Cliffs, New Jersey, 1967.

Acknowledgments. The second author is very much indebted to IBM Belgium and NFWO foundation for supporting his stay at the University of Colorado at Boulder which made the work on this and other papers possible.