RAIRO. INFORMATIQUE THÉORIQUE

GHEORGHE PĂUN

Some consequences of a result of Ehrenfeucht and Rozenberg

RAIRO. Informatique théorique, tome 14, nº 1 (1980), p. 119-122

<http://www.numdam.org/item?id=ITA_1980__14_1_119_0>

R.A.I.R.O. Informatique théorique/Theoretical Informatics
(vol. 14, n° 1, 1980, p. 119 à 122)

SOME CONSEQUENCES OF A RESULT OF EHRENFEUCHT AND ROZENBERG (*)

by Gheorghe PĂUN

Communiqué par J. BERSTEL

Abstract. — *Recently, Ehrenfeucht and Rozenberg have proved that there are context-free languages that are not EDTOL languages. In the present note, some consequences of this result are pointed out with reference to certain open problems related to matrix languages.*

Résumé. — *Récemment, Ehrenfeucht et Rozenberg ont prouvé qu'il existe des langages indépendants du contexte qui ne sont pas des EDTOL-langages. Dans la présente note, nous indiquons quelques conséquences de ce résultat, en résolvant certains problèmes ouverts sur les langages matriciels.*

1. NOTATIONS

In order to save space we shall omit the definitions and we shall specify only the notations we use:

— $\mathscr{L}$ (EDTOL) is the family of languages generated by EDTOL systems (*see* [6]);

— $\mathscr{M}_f$ is the family of languages generated by matrix grammars of finite index [1];

— $\mathscr{S}\mathscr{M}$ is the family of simple matrix languages [7] and $\mathscr{S}\mathscr{M}_f$ is the family of finite index languages in $\mathscr{S}\mathscr{M}$;

— $\mathscr{L}_2$ is the family of context-free languages;

— $\text{Ind}_m(L)$, $\text{Ind}_{cf}(L)$ denote the index of a language L according to matrix grammars, respectively, to context-free grammars;

— D_i is the Dyck language over the vocabulary $\{a_1, \ldots, a_i, a'_1, \ldots, a'_i\}$, that is the language generated by the context-free grammar with the rules $S \to SS$, $S \to a_j S a'_j$, $j=1, 2, \ldots, i$, $S \to \lambda$ [16].

(*) Reçu en mai 1979, et dans sa version définitive en juillet 1979.

(1) University of Bucharest, Division of Systems Studies, Bucuresti, Romania.

R.A.I.R.O. Informatique théorique/Theoretical Informatics, 0399-0540/1980/119/$ 5.00

2. OPEN PROBLEMS RELATED TO THE FAMILIES $\mathscr{M}_f$, $\mathscr{S}\mathscr{M}$

The index of a matrix grammar was defined in [1] where it is proved that for any infinite language L in $\mathscr{M}_f$, the set of lengths of strings in L contains an infinite arithmetical progression (a similar result holds for languages in $\mathscr{L}_2$). In [1] it is asked (problem A) whether or not any context-free language has a finite index (clearly, according to matrix grammars).

The problem was partially solved by Salomaa [15] by proving that $\mathrm{Ind}_{cf}(D_1)$ is infinite. The existence of a context-free language of infinite index according to matrix grammars remained open.

In [3], Cremers, Mayer, Weiss claim that $\mathrm{Ind}_m(D_1)=2$, but, as in [8] it was pointed out, the proof in [3] is wrong; the finiteness of $\mathrm{Ind}_m(D_1)$ remained open. Note that if any Dyck language would be in $\mathscr{M}_f$, then $\mathscr{L}_2 \subset \mathscr{M}_f$ since the family $\mathscr{M}_f$ is a full-AFL [11] and any language L in $\mathscr{L}_2$ can be written as $h(D_i \cap R)$, where h is a homomorphism and R is a regular language [16].

In [7] Ibarra introduces the so-called simple matrix grammars, a class of grammars generating languages with many context-free-like properties. Similar properties (closure properties, decision results, pumping lemmas) were proved for the family $\mathscr{M}_f$ too (*see* [11, 12, 14]). The relation between the families $\mathscr{M}_f$ and $\mathscr{S}\mathscr{M}$ was formulated as an open problem in various contexts [9, 11, 12]; in fact, we obviously have $\mathscr{M}_f - \mathscr{S}\mathscr{M} = \emptyset$ since the language $\{a^n b^n c^n \mid n \geqq 1\}^*$ is in $\mathscr{M}_f$ but not in $\mathscr{S}\mathscr{M}$ [7]. Therefore, the problem asks for a language in $\mathscr{S}\mathscr{M}$ which has an infinite matrix index. A stronger formulation [12] asks for context-free languages which are not in $\mathscr{M}_f$ (problem related to the paper [3]).

In [9] we proved that $\mathscr{S}\mathscr{M}_f \subset \mathscr{M}_f$ and one asks whether or not $\mathscr{S}\mathscr{M}_f \subseteqq \mathscr{S}\mathscr{M}$ is a proper inclusion. The problem was solved in [10] by proving that D_1 does not belong to $\mathscr{S}\mathscr{M}_f$. In a similar way it follows that any D_i, $i \geqq 1$, is not in $\mathscr{S}\mathscr{M}_f$.

3. THE SETTLEMENT OF THE ABOVE OPEN PROBLEMS

The result in [4] can be easily used to solve all the above open problems. Indeed, we have:

THEOREM: $\mathrm{Ind}_m(D_i)$ *is infinite for any* $i \geqq 2$.

Proof: In [2] one introduces the "parallel matrix grammars" as usual context-free matrix grammars with parallel derivation (each rule rewrites all occurrences of its left-hand side in the string to be rewritten). Let us denote by $\mathscr{P}\mathscr{M}$ the family of languages generated by these grammars. In [2] it is proved that $\mathscr{M}_f \subset \mathscr{P}\mathscr{M}$, strict inclusion.

In [13] it is proved that, in fact, we have $\mathcal{PM}=\mathcal{L}$ (EDTOL), therefore $\mathcal{M}_f \subset \mathcal{L}$ (EDTOL). Following [4], there are context-free languages which are not in $\mathcal{L}$ (EDTOL). Thus there are context-free languages which do not belong to $\mathcal{M}_f$ as well. Consequently, any context-free generator (any context-free language whose smallest AFL containing it equals the family of context-free languages) is not a matrix language of finite index. Any language D_i, $i \geqq 2$, is a context-free generator (*see* example 5.1.1, p. 139 [5]), hence $\mathrm{Ind}_m(D_i)$ is infinite for any $i \geqq 2$.

Consequently, (1) $\mathcal{L}_2$ and $\mathcal{M}_f$ are incomparable (clearly, $\mathcal{M}_f$ contains non-context-free languages), (2) $\mathcal{SM}$ and $\mathcal{M}_f$ are incomparable, (3) $\mathrm{Ind}_m(D_i)=\mathrm{Ind}_{cf}(D_i)=\infty$, $i \geqq 2$, and thus all the problems listed in section 2 are settled.

It remains open the problem to find context-free languages L such that $\mathrm{Ind}_m(L) < \mathrm{Ind}_{cf}(L)$ [or even $\mathrm{Ind}_m(L) < \mathrm{Ind}_{cf}(L)=\infty$]. Particularly, we do not know whether $\mathrm{Ind}_m(D_1)$ is finite or not.

Some problems settled by the result in [4] were considered in [4] too. Perhaps there are other problems which can be solved using the same result.

Note: We are very indebted to the referee for pointing out a mistake in an earlier version of this paper.

REFERENCES

1. B. Brainerd, *An Analog of a Theorem About Context-Free Languages*, Inform. Control, Vol. 11, 1968, pp. 561-568.
2. S. Ciobotaru, *The Parallel Matrix Grammars and the Thue Languages*, Bull. Math. Soc. Sc. Math. R.S.R., Vol. 27, (70), 1978, pp. 269-278.
3. A. B. Cremers, O. Mayèr and K. Weiss, *On the Complexity of Regulated Context-Free Rewriting*, Intern. Bericht, No. 4, Univ. Karlsruhe, Facultät für Informatik, 1973.
4. A. Ehrenfeucht and G. Rozenberg, *On Some Context-Free Languages that Are Not Deterministic ETOL Languages*, R.A.I.R.O./Theoretical Computer Sc., Vol. 11, No. 4, 1977, pp. 273-291.
5. S. Ginsburg, *Algebraic and Automata-Theoretic Properties of Formal Languages*, North-Holl. Publ. Comp., Amsterdam, Oxford, 1975.
6. G. Herman and G. Rozenberg, *Developmental Systems and Languages*, North-Holl. Publ. Comp., Amsterdam, 1975.
7. O. Ibarra, *Simple Matrix Languages*, Inform. Control, Vol. 17, 1970, pp. 359-394.
8. Gh. Păun, *On the Index of Grammars and Languages*, Inform. Control, Vol. 35, 1977, pp. 259-266.
9. Gh. Păun, *On the Generative Capacity of Simple Matrix Grammars of Finite Index*, Inform. Processing Letters, Vol. 7, 1978, pp. 100-102.

10. Gh. Păun, *On the Generative Capacity of Some Classes of Grammars with Regulated Rewriting*, Math. Found. of Computer Sc. Symp., Olomouc, Czechoslovakia, 1979.

11. Gh. Păun, *On the Family of Finite Index Matrix Languages*, J. Computer and Syst. Sc., Vol. 18, 1979, pp. 267-280.

12. Gh. Păun, *Matrix Grammars*, Editura Stiintificâ si Enciclopedicâ, Bucuresti, 1980 (in Romanian).

13. Gh. Păun, *On the Generative Capacity of Parallel Matrix Grammars*, Bull. Math. Soc. Sc. Math. R.S.R., Vol. 23, (71), 1979, pp. 367-371.

14. Gh. Păun, *Some Context-Free-Like Properties of Matrix Languages of Finite Index*, Bull. Math. Soc. Sc. Math. R.S.R. (in press).

15. A. Salomaa, *On the Index of Context-Free Languages*, Inform. Control, Vol. 14, 1969, pp. 474-477.

16. A. Salomaa, *Formal Languages*, Academic Press, New York and London, 1973.