JOURNAL OF COMPUTER AND SYSTEM SCIENCES **29**, 54–57 (1984)

Commutative One-Counter Languages Are Regular

M. LATTEUX

Université de Lille I, U.E.R. d'I.E.E.A.–Informatique, 59655 Villeneuve d'Ascq Cedex, France

AND

G. ROZENBERG

Institute of Applied Mathematics and Computer Science, University of Leiden, 2300 Ra Leiden, The Netherlands

Received February 1, 1983; revised September 10, 1983

A new characterization of commutative regular languages is given. Using it, it is proved that every commutative one-counter language is regular.

INTRODUCTION

The family of regular languages is a very basic and quite well understood family of languages. Hence, one way of gaining more insight into various language families is to investigate conditions which imposed on a language within a given language family will imply its regularity. This line of research is quite popular within formal language theory (see, e.g., [1, 3, 4] and [5]).

In the present paper we investigate conditions enforcing regularity of commutative languages. In particular we are concerned with those subfamilies of the family of context-free languages for which commutativity implies regularity.

It is conjectured in [8] that the family of quasi-rational languages (that is, the substitution closure of the family of linear languages) is a language family of such a kind. The conjecture still remains open, however, two partial answers have been recently established: every commutative linear language is regular (see [4]) and every commutative quasi-rational language over a two-letter alphabet is regular (see [10]). It is conjectured in [9] that commutativity implies regularity within the family of one-counter languages, that is, the rational cone generated by $D_1'^*$—the semi-Dyck language over one pair of parentheses (a rational cone is a family of languages closed under rational transductions or, equivalently, closed under morphisms, inverse morphisms and intersection with regular sets).

This conjecture is proved in our paper. In order to prove it, we provide a new characterization of commutative regular languages: a commutative language $L \subseteq X^*$

0022-0000/84 $3.00

is regular if and only if there exists a positive integer N such that, for each $w \in L$ and each $x \in X$, if the number of occurrences of x in w is at least N, then $w(x^N)^* \subseteq L$.

This characterization combined with a pumping lemma for one-counter languages provides a proof of the conjecture.

Preliminaries

We assume the reader to be familiar with the basic formal language theory, and in particular with the theory of context-free languages (see, e.g., [2, 6] and [11]).

Let X be a finite alphabet. For $w \in X^*$, $|w|$ denotes the length of w and for $x \in X$, $|w|_x$ denotes the number of occurrences of x in w. The (binary) shuffle operation ($\sqcup\!\sqcup$) on X^* is defined by $u \sqcup\!\sqcup v = \{u_1 v_1 u_2 v_2 \cdots u_n v_n \mid n \geqslant 1, u_i, v_i \in X^*, u_1 \cdots u_n = u$ and $v_1 \cdots v_n = v\}$.

For languages $L, L' \subseteq X^*$, $L \sqcup\!\sqcup L' = \bigcup_{w \in L, w' \in L'} w \sqcup\!\sqcup w'$.

The commutative closure of $w \in X^*$, denoted $c(w)$, is the set of all words obtained from w by permuting occurrences of letters in w. For a language $L \subseteq X^*$, $c(L) = \bigcup_{w \in L} c(w)$ and we say that L is a commutative language if $L = c(L)$.

I. Commutative Regular Languages

In this section we provide a characterization of commutative regular languages.

By using the shuffle operation, the celebrated theorem of Higman [7] can be stated in the following manner:

THEOREM 1 [7]. *For every language $L \subseteq X^*$, there exists a finite subset F of L such that $L \subseteq F \sqcup\!\sqcup X^*$. Hence, $L \sqcup\!\sqcup X^* = F \sqcup\!\sqcup X^*$ is a regular language.*

We will need a slight generalization of this result.

THEOREM 2. *Let $X = \{x_1, \ldots, x_k\}$ be an alphabet and let for each $i \in \{1, \ldots, k\}$ u_i be a word in x_i^+. Then, for each language $L \subseteq X^*$, $L_1 = L \sqcup\!\sqcup c(u_1^* \cdots u_k^*)$ is a regular language.*

Proof. Let $A = \{(i_1, \ldots, i_k) \mid 0 \leqslant i_j < |u_j|$ for $j \in \{1, \ldots, k\}\}$ and let, for $\rho = (i_1, \ldots, i_k) \in A$, R_ρ be the regular language $c(x_1^{i_1} u_1^* \cdots x_k^{i_k} u_k^*)$. Clearly $L_1 = \bigcup_{\rho \in A} (((L \cap R_\rho) \sqcup\!\sqcup X^*) \cap R_\rho)$. Since A is finite and (by Theorem 1) $(L \cap R_\rho) \sqcup\!\sqcup X^*$ is regular, L_1 is regular. ∎

We are ready now to prove a characterization of commutative regular languages.

THEOREM 3. *Let $L \subseteq X^*$ be a commutative language. Then L is regular if and only if there exists a positive integer N such that for each $w \in L$ and for each $x \in X$, $|w|_x \geqslant N$ implies $w(x^N)^* \subseteq L$.*

Proof. Let $X = \{x_1, ..., x_k\}$. Assume that L is regular. Then $L \cap x_1^* \cdots x_k^*$ is a finite union of languages of the form $A_1 A_2 \cdots A_k$, where, for $1 \leqslant i \leqslant k$, A_i is a regular subset of x_i^*. Since a regular language included in x_i^* is a finite union of languages of the form $x_i^p(x_i^q)^*$, we get $L \cap x_1^* \cdots x_k^* = \bigcup_{j=1}^n L_j$, where, for $1 \leqslant j \leqslant n$, $L_j = x_1^{s_1}(x_1^{t_1})^* \cdots x_k^{s_k}(x_k^{t_k})^*$ for some $s_1, t_1, ..., s_k, t_k \in \mathbb{N}$. Let, for $1 \leqslant j \leqslant n$, d_j denote $\sup\{s_1 + t_1, ..., s_k + t_k\}$, $d = 1 + \sup\{d_j \mid 1 \leqslant j \leqslant n\}$ and let $N = d!$.

Let $i \in \{1, ..., k\}$ and let $w \in L$ be such that $|w|_{x_i} \geqslant N$. Then there exists $L_j = x_1^{s_1}(x_1^{t_1})^* \cdots x_k^{s_k}(x_k^{t_k})^*$ such that $w \in c(L_j)$ and $1 \leqslant t_i \leqslant d$. Clearly, $w(x_i^{t_i})^* \subseteq c(L_j) \subseteq L$ and consequently (because t_i divides N) $w(x_i^N)^* \subseteq L$.

Assume now that L satisfies the sufficient part of the statement of the theorem. Then the regularity of L is proved by the induction on k—the number of distinct letters occurring in the words from L.

If $k = 0$, then $L = \{\varepsilon\}$ is regular.

Assume that $k > 0$. Then $L = L' \cup L''$, where $L' = \{w \in L \mid |w|_x \geqslant N$ for each $x \in X\}$ and $L'' = \bigcup_{i=1}^k (\bigcup_{j=0}^{N-1} L_{i,j})$ for $L_{i,j} = \{w \in L \mid |w|_{x_i} = j\}$. For $i \in \{1, ..., k\}$, let $X_i = X \setminus \{x_i\}$ and let $\underline{\text{pres}}_{X_i}: X^* \to X_i^*$ be the homomorphism defined by $\underline{\text{pres}}_{X_i}(x) = x$ for $x \in X_i$ and $\underline{\text{pres}}_{X_i}(x_i) = \varepsilon$. Then $L_{i,j} = L'_{i,j} \sqcup\!\sqcup\, x_i^j$, where $L'_{i,j} = \underline{\text{pres}}_{X_i}(L_{i,j})$ is a commutative language over X_i. Let $w \varepsilon L'_{i,j}$ and $x_s \in X_i$ be such that $|w|_{x_s} \geqslant N$. Then $wx_i^j \in L_{i,j} \subseteq L$ and so, by assumption, $wx_i^j(x_s^N)^* \subseteq L$. Since $wx_i^j \in L_{i,j}$ and $s \neq i$, it must be that $wx_i^j(x_s^N)^* \subseteq L_{i,j}$ and consequently $w(x_s^N)^* \subseteq L'_{i,j}$. By the inductive hypothesis it follows that $L'_{i,j}$ is regular and consequently $L_{i,j}$ and L'' are regular.

Hence it remains to prove that L' is regular. By the definition of L', we have $L'(x_1^N)^* \cdots (x_k^N)^* \subseteq L'$ which implies that $L' = L'(x_1^N)^* \cdots (x_k^N)^*$ and consequently (because L' is a commutative language) $L' = L' \sqcup\!\sqcup\, c((x_1^N)^* \cdots (x_k^N)^*)$. Thus, by Theorem 2, L' is regular. ∎

II. Commutative One-Counter Languages

In this section we prove that every commutative one-counter language is regular. First we recall an iteration result for one-counter languages.

THEOREM 4 [9]. *Let L be a one-counter language. There exists a positive integer N_0 such that each $w \in L$ with $|w| \geqslant N_0$ admits a factorization $w = w_1 u w_2 v w_3$ satisfying the following conditions*:

$|uv| \geqslant 1$,

$|w_1 u v w_3| \leqslant N_0$,

$w_1 u^n w_2 v^n w_3 \in L$ *for each* $n \geqslant 1$.

THEOREM 5. *Every commutative one-counter language is regular.*

Proof. Let $L \subseteq X^*$ be a commutative one-counter language and let $N_0 \geqslant 3$ be a constant satisfying the statement of Theorem 4. Let $N = N_0! \geqslant 2N_0$. Let $w \in L$ and

$x \in X$ be such that $|w|_x \geqslant N$. Then there exists a word $w' \in c(w) \subseteq L$ such that $w' = x^{N_0} w'' x^{N_0}$ for some $w'' \in X^*$. By Theorem 4, w' admits a factorization $w' = w_1 x^i w_2 x^j w_3$, where $1 \leqslant i+j \leqslant N_0$ and $w_1(x^i)^n w_2(x^j)^n w_3 \in L$ for each $n \geqslant 1$. Since $w \in c(w')$ and L is commutative, $w(x^{i+j})^* \subseteq L$; consequently, because $i+j$ divides N, $w(x^N)^* \subseteq L$. Hence, by Theorem 3, L is regular and the theorem holds. ∎

Inspecting the proof of the above result we notice that actually one has the following result: if $\mathscr{L}$ is a language family for which the iterative result of Theorem 4 holds, then every commutative language in $\mathscr{L}$ is regular. But this iteration theorem is weaker than the classical iteration lemma for linear languages (see, e.g., [2]) and so we get a different proof of the result from [4] which states that every commutative linear language is regular.

References

1. J. M. Autebert, J. Beauquier, L. Boasson, and M. Latteux, Very small families of algebraic non-rational languages, *in* "Formal Language Theory" (R. Book, Ed.), pp. 89–108, Academic Press, New York, 1980.
2. J. Berstel, "Transductions and Context-Free Languages," Teubner, Stuttgart, 1979.
3. L. Boasson, Un critère de rationalité des languages algébriques, *in* "Automata, Languages and Programming" (M. Nivat, Ed.), pp. 359–365, North-Holland, Amsterdam, 1973.
4. A. Ehrenfeucht, D. Haussler, and G. Rozenberg, Conditions enforcing regularity of context-free languages, 9th ICALP, Aarhus, 1982, *in* "Lecture Notes in Computer Science," Vol. 140, pp. 187–191.
5. A. Ehrenfeucht, R. Parikh, and G. Rozenberg, Pumping lemmas for regular sets, *SIAM J. Comput.* **10** (1981), 536–541.
6. M. Harrison, "Introduction to Formal Language Theory," Addison–Wesley, Reading, Mass., 1978.
7. G. H. Higman, Ordering by divisibility in abstract algebras, *Proc. London Math. Soc.* **3** (1952), 326–336.
8. M. Latteux, "Langages commutatifs," Thèse Sc. Math., Lille I, 1978.
9. M. Latteux, Langages à un compteur, *J. Comput. Systems Sci.* **26** (1983), 14–33.
10. M. Latteux and J. Leguy, On the usefulness of bifaithful rational cones, Second Conference on Foundations of Software Technology and Theoretical Computer Science, Bangalore, 1982.
11. A. Salomaa, "Formal Languages," Academic Press, New York, 1973.