Power-Separating Regular Languages

by

H. J. SHYR* and G. THIERRIN*

Department of Mathematics
University of Western Ontario
London, Ontario, Canada

1. Introduction. We call a language U a *power-separating* language if there is a positive integer $m \geq 1$ such that for every word x either $\{x^m, x^{m+1}, \cdots\} \cap U = \varnothing$ or $\{x^m, x^{m+1}, \cdots\} \subseteq U$. In this paper we determine some properties of the power-separating regular languages and their relation with the noncounting regular languages. In general the class of noncounting regular languages is properly contained in the class of power-separating regular languages. The automata accepting such languages are discussed and we establish a decomposition of regular languages in function of power-separating regular languages.

2. Definitions and Properties of Power-Separating Regular Languages. Throughout this paper X will represent a finite alphabet and X^* the free semigroup generated by X. Any subset U of X^* will be called a language over X. A language U is called regular if U is accepted by a finite automaton; this is equivalent to the property that U is a union of some of the equivalence classes of a right congruence relation over X^* of finite index. The symbol $| \ |$ will stand for the cardinality of a set. We recall the following definition of a noncounting language (see [4]).

Definition 1. A language $U \subseteq X^*$ is a *noncounting language* over X if and only if there is a non-negative integer k (k dependent on U) such that for all $x, y, z \in X^*$, $xy^kz \in U \Leftrightarrow xy^{k+1}z \in U$.

Definition 2. A word $x \in X^*$ is called *power-free* if and only if $v^n = x$ implies $n = 1$, where $v \in X^*$. The set of all power-free words over X will be denoted by $P(X)$. For any $x \in X^*$ and $m \geq 0$, let $J_x^m = \{x^n | n \geq m\}$. In particular, $J_x^0 = \{\Lambda, x, x^2, \cdots\} = x^*$.

Definition 3. (1) A language U over X is called a *quasi-power-separating language* or simply *qp-separating language* if and only if for every $x \in X^*$ there is an integer $m(x) \geq 1$ such that either $J_x^{m(x)} \cap U = \varnothing$ or $J_x^{m(x)} \subseteq U$. The minimal

* This research has been supported by Grant A7877 of the National Research Council of Canada.

positive integer $m(x)$ satisfying the above property will be called the *height* of x relatively to U.

(2) A language U over X is called a *power-separating language* or simply *p-separating language* if and only if there is an integer $m \geq 1$ such that for all $x \in X^*$ either $J_x^m \cap U = \varnothing$ or $J_x^m \subseteq U$. The minimal positive integer m satisfying the above property will be called the height of U.

By a (*qp*) *p*-separating regular language we shall mean a (*qp*) *p*-separating language which is regular.

We observe that if $x \in X^*$ is a power-free word and $v = x^n$ for some $n \geq 1$, $v \in X^*$, then $J_v^m \subseteq J_x^m$ for every integer $m \geq 1$. The following theorem is now obvious.

PROPOSITION 1. *(1) A language U over X is a qp-separating language if and only if for every $x \in P(X)$ there is an integer $m(x) \geq 1$ such that either $J_x^{m(x)} \cap U = \varnothing$ or $J_x^{m(x)} \subseteq U$. (2) A language U over X is a p-separating language if and only if there is an integer $m \geq 1$ such that for every $x \in P(X)$ either $J_x^m \cap U = \varnothing$ or $J_x^m \subseteq U$.*

PROPOSITION 2. *Let $x \in X^*$, $x \neq \Lambda$. Then $x \in P(X)$ if and only if J_x^1 is a p-separating regular language.*

Proof. Suppose $x \in X^*$ and J_x^1 is a *p-separating* regular language; then x is power-free. Indeed, if x is not a power-free word, say $x = v^n$, $n \neq 1$, for some $v \in X^*$, then for every $k \geq 1$, $v^{nk} \in J_x^1$ but $v^{nk+1} \notin J_x^1$.

Conversely, assume $x \in P(X)$; then, since $J_x^1 = x^* \setminus \{\Lambda\}$, J_x^1 is a regular language. Now suppose $v \in P(X)$. We will show that $J_v^1 \cap J_x^1 = \varnothing$ or $J_v^1 \subseteq J_x^1$. Let us suppose that $J_v^1 \cap J_x^1 \neq \varnothing$. Then there exist $k \geq 1$ and $r \geq 1$ such that $v^k = x^r$. Then, by Lemma 5.5.2 ([2]), $vx = xv$. Let $Y = \{v, x\}$; then Y^* is a commutative set such that $Y^* \subseteq X^*$ (commutative set means every pair of words commute (see [2])). By Lemma 5.5.1 ([2]), there is a word ω such that $Y^* \subseteq \omega^*$ and so $v = \omega^{h_1}$, $x = \omega^{h_2}$ for some $h_1, h_2 \geq 1$. It follows that $h_1 = h_2$ and $v = x$ holds, since x, v are power-free. Thus $J_v^1 = J_x^1$. This completes the proof of the proposition.

PROPOSITION 3. *Every noncounting regular language U over X is a p-separating regular language.*

Proof. Suppose U is a noncounting regular language. Then by definition there exists an integer $k \geq 0$ such that, for every $x, y, z \in X^*$, $xy^kz \in U \Leftrightarrow xy^{k+1}z \in U$. Let $m = k+1$ and $v \in X^*$. Then, if $v^m \notin U$ then $v^{m+r} \notin U$ for all $r \geq 1$ and on the other hand, if $v^m \in U$ then $v^{m+r} \in U$ for all $r \geq 1$. Thus U is a *p*-separating regular language.

PROPOSITION 4. *Let $X = \{x\}$ and $U \subseteq X^*$. Then the following are equivalent. (1) U is a noncounting regular language; (2) U is a p-separating regular language.*

Proof. (1) $\Rightarrow$ (2) follows directly from Proposition 3. Now we prove (2) $\Rightarrow$ (1). Suppose $U \subseteq X^*$ is a *p*-separating regular language. Then by definition there exists $m \geq 1$ such that either $J_x^m \cap U = \varnothing$ or $J_x^m \subseteq U$. If $J_x^m \cap U = \varnothing$ then U is finite and so U is a noncounting regular language. If on the other hand $J_x^m \subseteq U$, then clearly U is a noncounting regular language. Thus (2) $\Rightarrow$ (1).

In general, the class of p-separating regular languages contains the class of noncounting regular languages properly. The following example will show this fact.

Example 1. Let $X = \{a, b\}$. The regular language $L = \{a^{2m}b^{2n} | m, n \geq 1\}$ is not a noncounting regular language. For if k is even, then $a^2a^kb^2 \in L$ but $a^2a^{k+1}b^2 \notin L$; if k is odd, then $aa^kb^2 \in L$ but $aa^{k+1}b^2 \notin L$. L is a p-separating regular language. Indeed, if we take $m = 2$, then for every $x \in X^*$, $J_x^2 \cap L = \varnothing$.

3. Operations on p-Separating Regular Languages.

PROPOSITION 5. *The class of p-separating regular languages over X contains X^* and all the finite subsets of X^* and it is closed under the Boolean operations of union, intersection and complementation.*

Proof. It is clear that X^* is a p-separating regular language. If U is a finite subset of X^* and the maximal length of words in U is m, then for every $x \neq \Lambda$, $x \in P(X)$, we have $J_x^{m+1} \cap U = \varnothing$. Thus every finite subset of X^* is a p-separating regular language.

Let us prove now that this class is closed under Boolean operations. Suppose that U_1 and U_2 are in the class with the heights m_1 and m_2 respectively. We have for any $x \in X^*$, either $J_x^{m_1} \cap U_1 = \varnothing$ or $J_x^{m_1} \subseteq U$ and either $J_x^{m_2} \cap U = \varnothing$ or $J_x^{m_2} \subseteq U_2$. Take $m = m_1 + m_2$. It is easy to see that, for any $x \in X^*$, either $J_x^m \cap (U_1 \cap U_2) = \varnothing$ or $J_x^m \subseteq (U_1 \cap U_2)$ and that either $J_x^m \cap (U_1 \setminus U_2) = \varnothing$ or $J_x^m \subseteq (U_1 \setminus U_2)$. Thus the class of p-separating regular languages is closed under the Boolean operations $\cup$, $\cap$ and $-$.

In general the concatenation of two p-separating regular languages may not be a p-separating regular language. For example, $A = \{(ab)^{2^n}a | n \geq 1\}$ and $B = \{b(ab)^{2m+1} | m \geq 1\}$ are two p-separating regular languages. However, the concatenation $AB = \{(ab)^{2n} | n \geq 3\}$ is regular but not p-separating.

For a p-separating regular language U, the sets $U .\cdot d = \{x \in X^* | dx \in U\}$ and $U \cdot. d = \{x \in X^* | xd \in U\}$ are in general not p-separating languages. For example, in Example 1, the language $L = \{a^{2m}b^{2n} | m, n \geq 1\}$ is a p-separating regular language over $X = \{a, b\}$ but both $L \cdot. a^2b^2 = \{a^{2m} | m \geq 0\}$; $L .\cdot a^2b^2 = \{b^{2n} | n \geq 0\}$ are not p-separating languages.

A p-separating language may fail to be a regular language. The following proposition will show this fact.

PROPOSITION 6. *Let $P(X)$ be the set of all power-free words over X. Then the following are true. (1) $P(X)$ is a p-separating language; (2) If $|X| = 1$, then $P(X)$ is regular; (3) If $|X| \geq 2$, then $P(X)$ is not regular.*

Proof. (1) If $v \in P(X)$, then $J_v^2 \cap P(X) = \varnothing$. Hence $P(X)$ is a p-separating language.

(2) If $|X| = 1$, then $|P(X)| < \infty$. Hence $P(X)$ is regular.

(3) Assume $X = \{a, b, \cdots, d\}$ and $|X| \geq 2$. If $P(X)$ is a union of some classes of a congruence relation, then the cardinality of the congruence classes is not finite. To show this, let $[x]$ denote the class containing x and observe that $[a]$, $[b]$ and $[ab]$ form different classes. Let us prove now by induction on k that for any $k \geq 1$, $[a]$, $[b]$, $[ab]$, $[a^2b^2], \cdots, [a^kb^k]$ form different classes. Assume

the proposition is true for k. Consider the word $a^{k+1}b^{k+1} \in P(X)$. It is sufficient to show that $a^{k+1}b^{k+1}$ does not belong to the classes $[a]$, $[b]$, $[ab]$, $[a^2b^2], \cdots,$ $[a^kb^k]$. This is true, because $a^2 \notin P(X)$ and $a^{k+2}b^{k+1} \in P(X)$, $b^2 \notin P(X)$ and $b^{k+1}b^{k+2} \in P(X)$, $(a^ib^i)(a^ib^i) \notin P(X)$ and $a^{k+1}b^{k+1}a^ib^i \in P(X)$ for $i = 1, 2, \cdots, k$. Thus $a^{k+1}b^{k+1}$ belongs to a new congruence class. Hence the cardinality of the congruence classes is infinite and (3) follows.

4. Automata Accepting p-Separating Regular Languages. Let T be a monoid and let H be a subset of T. The relation P_H defined on T by $aP_Hb \Leftrightarrow H..a = H..b$, where $H..a = \{(x, y)|x, y \in T, xay \in H\}$ is a congruence of T called the principal congruence determined by H (see [1]), and H, if not empty, is a union of classes of P_H. The set H is said to be disjunctive if and only if P_H is the equality.

If U is a language over the alphabet X, then the quotient monoid X^*/P_U is called the syntactic monoid of U and it is denoted by Syn (U). It is well known that U is a regular language if and only if P_U is of finite index, that is if and only if Syn (U) is finite.

Definition 4. Let $A = (S, X, M, s_1, F)$ be an automaton. (1) A is called *qp-separating* if and only if for every $x \in X^*$, there is an integer $m(x) \geq 1$ such that either $\{M(s_1, x^k)|k \geq m(x)\} \cap F = \varnothing$ or $\{M(s_1, x^k)|k \geq m(x)\} \subseteq F$. (2) A is called *p-separating* if and only if there is an integer $m \geq 1$ such that for every $x \in X^*$ either $\{M(s_1, x^k)|k \geq m\} \cap F = \varnothing$ or $\{M(s_1, x^k)|k \geq m\} \subseteq F$.

PROPOSITION 7. *Let U be a regular language over X. Then the following are equivalent. (1) U is accepted by a qp-separating automaton; (1)′ U is accepted by a p-separating automaton; (2) U is a qp-separating regular language; (2)′ U is a p-separating regular language; (3) P_U is a congruence of finite index and for every $x \in X^*$ there exists an integer $m(x) \geq 1$ such that either $J_x^{m(x)} \cap U = \varnothing$ or $J_x^{m(x)} \subseteq U$; (3)′ P_U is a congruence of finite index and there exists an integer $m \geq 1$ such that for every $x \in X^*$ either $J_x^m \cap U = \varnothing$ or $J_x^m \subseteq U$.*

Proof. First we show (1) $\Leftrightarrow$ (1)′. That (1)′ $\Rightarrow$ (1) is clear. Suppose (1) holds. Then there exists an automaton $A = (S, X, M, s_1, F)$ such that for any $x \in X^*$ there is an integer $m(x) \geq 1$ such that either $\{M(s_1, x^n)|n \geq m(x)\} \cap F = \varnothing$ or $\{M(s_1, x^n)|n \geq m(x)\} \subseteq F$ and $U = T(A)$. For every x, let $m_0(x) \geq 1$ be the minimal positive integer satisfying the above condition. Since S is finite, the function $m_0(x)$ is such that $|\{m_0(x)|x \in X^*\}| < \infty$ holds. Let $m = \max \{m_0(x)| x \in X^*\}$. Then for every $x \in X^*$ either $\{M(s_1, x^n)|n \geq m\} \cap F = \varnothing$ or $\{M(s_1, x^n)|n \geq m\} \subseteq F$. This shows that A is a p-separating automaton and $U = T(A)$. Thus (1) $\Rightarrow$ (1)′.

Next we show (1)′ $\Rightarrow$ (2)′ $\Rightarrow$ (3)′ $\Rightarrow$ (1)′.

Suppose (1)′ holds, that is, there is a p-separating automaton $A = (S, X, M, s_1, F)$ accepting U. Then by definition there is an integer $m \geq 1$ such that for every $x \in X^*$ either $\{M(s_1, x^n)|n \geq m\} \cap F = \varnothing$ or $\{M(s_1, x^n)|n \geq m\} \subseteq F$. It follows that for every $x \in X^*$ either $J_x^m \cap U = \varnothing$ or $J_x^m \subseteq U$. Thus U is a p-separating regular language and (1)′ $\Rightarrow$ (2)′ hold.

Now suppose (2)′. Since U is a regular language, (3)′ follows directly from (2)′.

To show (3)′ ⇒ (1)′, assume (3)′ and as usual we construct an automaton $A = (S, X, M, s_1, F)$ as follows: Take S to be the classes of P_U, $s_1 = [\Lambda]$ and F be the set of all $[x]$ where $x \in U$. The transition map is defined by $M([x], y) = [xy]$, for all $x, y \in X^*$. It is clear that $U = T(A)$ and for every $x \in X^*$, either $\{M(s_1, x^n) | n \geq m\} \cap F = \varnothing$ or $\{M(s_1, x^n) | n \geq m\} \subseteq F$, since by assumption for every $x \in X^*$, either $J_x^m \cap U = \varnothing$ or $J_x^m \subseteq U$. This completes the proof of (1)′ ⇔ (2)′ ⇔ (3)′.

Similarly, we can prove (1) ⇔ (2) ⇔ (3) by the same argument.

It is well known that if the regular language U is noncounting, then its syntactic monoid contains only trivial subgroups. This result can be generalized to the case of p-separating regular languages in the following way.

PROPOSITION 8. *Let U be a p-separating regular language over X and let* Syn (U) *be the syntactic monoid of U. Then* Syn (U) *contains a disjunctive set H such that for any subgroup G of* Syn (U), *$G \cap H \neq \varnothing$ implies $G \subseteq H$.*

Proof. Let $H = \{a | a \in \text{Syn}(U) \text{ and } a = [x], \text{ where } x \in U\}$. Since Syn $(U) = X^*/P_U$, it is immediate that H is a disjunctive subset of Syn (U) and, since U is regular, Syn (U) is a finite monoid.

If m is the height of U, then for every $x \in X^*$ either $J_x^m \cap U = \varnothing$ or $J_x^m \subseteq U$.

Let $a \in G \cap H$, where G is a subgroup of Syn (U) and let e be the identity of G. For every $b \in G$, we have $b^n = e$, where n is the order of G. Let $e = [z]$, where $z \in X^*$. If $e \notin H$, then for every integer $k \geq 1$, $e^k \notin H$ and therefore $z^k \notin U$. This implies that $J_z^m \cap U = \varnothing$. Let $a = [x]$, $x \in X^*$, if $J_x^m \cap U = \varnothing$, then $x^{mn+1} \notin U$ and $a = a^{mn+1} = [x^{mn+1}] \notin H$, a contradiction. Hence $J_x^m \subseteq U$, $x^{nm} \in U$ and $e = a^{nm} = [x^{nm}] \in H$.

Let $b = [y]$, $y \in X^*$. If $J_y^m \cap U = \varnothing$, then $y^{mn} \notin U$ and $e = b^{nm} = [y^{nm}] \notin H$, a contradiction. Hence $J_y^m \subseteq U$, $y^{nm+1} \in U$ and $b = b^{nm+1} = [y^{nm+1}] \in H$. Therefore $G \subseteq H$.

COROLLARY 1. *The only p-separating regular languages U over X such that* Syn (U) *is a group are X^* and $\varnothing$.*

5. A Decomposition of Regular Language Relatively to p-Separating Regular Languages. Following [10], a language $U \neq \varnothing$ is called right pure if $Ux \cap U \neq \varnothing$ implies $x = \Lambda$. If $\Lambda \notin U$, then U is right pure if and only if U is a prefix code ([9]). In [10], it has been shown that a language U is regular if and only if there exists a finite number of right pure languages $V_1, V_2, \cdots, V_k, W_1, W_2, \cdots, W_k$ such that $U = \bigcup_{i=1}^{k} V_i W_i^*$. We will have a similar representation relatively to the p-separating regular languages. First we prove the following proposition.

PROPOSITION 9. *Every right pure language U over X is a p-separating regular language.*

Proof. Let $x \in P(X)$. If $x \in U$, then clearly $x^n \notin U$ for every $n \geq 2$ and hence $J_x^2 \cap U = \varnothing$ is true. If $x \notin U$, then either $J_x^1 \cap U = \varnothing$ or there is an integer m such that $x^m \in U$. But the latter case will imply $J_x^{m+1} \cap U = \varnothing$. Hence by Proposition 1, U is a qp-separating language which is a p-separating language by Proposition 7. Thus every right pure language is a p-separating language.

PROPOSITION 10. *A language U is regular if and only if there exists a finite number of p-separating regular languages* $V_1, V_2, \cdots, V_k, W_1, W_2, \cdots, W_k$ *such that* $U = \bigcup_{i=1}^{k} V_i W_i^*$.

Proof. The proof is straightforward by applying the above proposition and the result mentioned above.

REFERENCES

[1] A. H. Clifford and G. B. Preston, *The Algebraic Theory of Semigroups*, Math. Surveys Series, Vols. 1 and 2, Amer. Math. Soc., Providence, R.I., 1961 and 1967.
[2] S. Ginsburg, *The Mathematical Theory of Context-Free Languages* McGraw-Hill, New York, 1968.
[3] A. Ginzburg, *Algebraic Theory of Automata*, Assoc. Comput Mach. Monograph Series, Academic Press, New York, 1968.
[4] R. McNaughton and S. Papert, *Counter-free Automata*, M.I.T. Press, 1971.
[5] A. R. Meyer, A note on star-free events, *J. Assoc. Comput. Mach.* **16** (1969), 220–225.
[6] M. O. Rabin and D. Scott, Finite automata and their decision problems, *IBM J. Res. Develop.* **3** (1959), 114–125.
[7] M. P. Schützenberger, On finite monoids having only trivial subgroups. *Information and Control* **8** (1965), 190–194.
[8] M. P. Schützenberger, On a family of sets related to McNaughton's L-language, In *Automata Theory*, E. R. Caianiello, (Ed.), Academic Press, New York, (1966), pp. 320–324.
[9] M. P. Schützenberger, On synchronizing prefix codes, *Information and Control* **11** (1967), 396–401.
[10] G. Thierrin, Décomposition des Langages Réguliers, R.I.R.O. 3e annee, N° R-3, (1969), 45–50.

(*Received 19 June 1972*)