ON REGULAR SEPARATION OF LANGUAGES

Arikawa, Setsuo
Research Institute of Fundamental Information Science, Kyushu University

https://doi.org/10.5109/13085

出版情報：統計数理研究. 16 (1/2), pp.83-94, 1974-03. Research Association of Statistical Sciences
バージョン：
権利関係：

ON REGULAR SEPARATION OF LANGUAGES

By

Setsuo ARIKAWA*

(Received September 30, 1973)

1. Introduction.

Languages $L_1, L_2, \cdots, L_n$ are said to be *regularly separable if there exist mutually disjoint regular sets* $R_1, R_2, \cdots, R_n$ *with* $L_i \subseteq R_i$ *for each* i. In case $n=2$, we say that L_1 is regularly separable from L_2, or regular set R_1 separates L_1 from L_2.

This notion, introduced by K. Kobayashi [1] in developing his abstract theory of complexity of formal languages, seems very important not only in the study of complexity but also in the general study of formal languages from the three points of views:

(1) It is useful for setting up a criterion of nearness of 'approximation' of languages by ones with a weaker structure, e. g., context-free languages by regular sets. When languages $L_1, L_2, \cdots, L_n$ are regularly separable, we can say that the regular superset R_1 of L_1 approximates L_1 as near as R_1 intersects with no other set L_i but L_1.

(2) It gives a negative way for language recognition. From the regular separability of languages L_1 and L_2, we can know, by only finite automata, to which language of L_1^c, L_2^c or $(L_1 \cup L_2)^c$ a given word belongs, though we can not have any positive answer.

(3) We may take, in a sense, that the regular sets in regular separation of languages correspond to the linear discriminant functions in the usual pattern recognition.

In the present paper, we shall study some aspects of regular separation of languages. In Section 2 we shall prove a fundamental lemma which gives a sufficient condition for regular separability of languages, by which we shall exhibit several examples of regularly separable languages, and give a necessary and sufficient condition for regular separability of languages from a set of all primes for the case of a singleton alphabet. In Section 3 we shall consider languages which are not regularly separable from any other infinite language, and show that the full set A^* of words can be covered by some finite union of context-sensitive languages which contain no infinite regular set, but not by any such context-free language. In the final section, we shall give a necessary and sufficient condition for regular sets containing no nonregular context-free language, and give a solvable decision problem.

* Research Institute of Fundamental Information Science, Kyushu University.

2. Regularly separable languages.

LEMMA 2.1. *If each of L_1 and L_2 is regularly separable from L, then so is each of $L_1 \cap L_2$ and $L_1 \cup L_2$.*

PROOF. Let a regular set R_1 separate L_1 from L, and a regular set R_2 separate L_2 from L. Then

$$L_1 \cap L_2 \subseteq L_1 \subseteq R_1 ,$$

$$L_1 \cup L_2 \subseteq R_1 \cup R_2 , \qquad L \subseteq R_1^c \cap R_2^c ,$$

$$(R_1 \cup R_2) \cap (R_1^c \cap R_2^c) = \phi ,$$

where L^c denotes the complement of L with respect to the full set.

Note that the regular separability are not always preserved under the operations of product and *. By the lemma and fundamental properties on regular sets, we may take mainly the case of $n=2$.

LEMMA 2.2. *Languages L_1 and L_2 are regularly separable if there exists a generalized sequential machine S for which $S(L_1)$ and $S(L_2)$ are mutually disjoint regular sets.*

PROOF. $S^{-1}(S(L_1))$ is a regular superset of L_1. And

$$\begin{aligned} S^{-1}(S(L_1)) \cap S^{-1}(S(L_2)) &= S^{-1}(S(L_1) \cap S(L_2)) \\ &= S^{-1}(\phi) = \phi . \end{aligned}$$

Hence the regular set $S^{-1}(S(L_1))$ separates L_1 from L_2.

EXAMPLE 2.1. *$L = \{a^n b^n ;\ n \geqq 0\}$ is regularly separable from $L' = \{a^n b^{n+2} ;\ n \geqq 0\}$.* In fact, we can construct a generalized sequential machine $S = (\{s_0, s_1, s_2\}, \{a, b\}, \{a, b, c\}, \delta, \lambda, s_0)$:

$$\delta(s_0, a) = s_1 , \qquad \delta(s_1, a) = s_2 , \qquad \delta(s_2, a) = s_0 ,$$

$$\delta(s_i, b) = s_i \qquad \text{for } i = 0, 1, 2 ;$$

$$\lambda(s_i, a) = \varepsilon \qquad \text{for } i = 0, 1, 2 ,$$

$$\lambda(s_0, b) = a , \qquad \lambda(s_1, b) = b , \qquad \lambda(s_2, b) = c .$$

And we have

$$S(L) = (a^3)^* \cup (b^3)^* b \cup (c^3)^* cc ,$$

$$S(L') = (a^3)^* aa \cup (b^3)^+ \cup (c^3)^* c ,$$

which are regular and mutually disjoint.

EXAMPLE 2.2. *$L = \{a^n b^n ;\ n \geqq 0\}$ is not regularly separable from $L' = \{a^n b^{2n} ;\ n > 0\}$.* Suppose that L is separated by a regular set R from L'. From the finiteness of the number of right invariant equivalence classes induced by R, we can see that there exists an integer m for which both a^m and a^{2m} are contained in one class, and hence there is a class which intersects with both L and L'.

EXAMPLE 2.3. *$L = \{a^{2^n} ;\ n \geqq 1\}$ and $L' = \{a^{3^n} ;\ n \geqq 1\}$ are regularly separable.*

EXAMPLE 2.4. *$L = \{a^{2^n} ;\ n \geqq 1\}^c$ and $L' = \{a^{4^n} ;\ n \geqq 1\}$ are not regularly separable,* for no set R with $L \subset R \subset L'^c$ is regular.

Here we note that the converse of Lemma 2.2 is not always true. In fact, for the regularly separable languages L and L' in the example 2.3, we can see that $S(L)\cap S(L')$ is not empty or not regular for any generalized sequential machine S.

Finally we give a necessary and sufficient condition for regular separability of languages over a singleton alphabet from the set of all primes.

THEOREM 2.1. *Let $P=\{a^p;\ p$ is a prime$\}$ and L be any language on a^* disjoint from P. Then L is regularly separable from P if and only if there exists a finite set F of integers greater than one such that $n=d\cdot l$ for any a^n in L with $n>1$ and for some d in F and $l(\geqq 2)$.*

PROOF. Suppose that there exists a finite set $F=\{d_1, d_2, \cdots, d_n\}$ which satisfies the condition. Then obviously

$$L\subseteqq \bigcup_{i=1}^{r}(a^{d_i})^*a^{2d_i}, \qquad P\cap\bigcup_{i=1}^{r}(a^{d_i})^*a^{2d_i}=\phi.$$

Conversely suppose that a regular set R separates P from L. Let $\{R_i\}_{i=1}^{k}$ be a set of right invariant equivalence classes induced by R. Every infinite R_j can be written as $R_j=a^{p_j}(a^{q_j})^*$, by which we have

$$P\cap R_j=\phi \Leftrightarrow p_j+q_j\cdot n \text{ is not prime for any } n$$
$$\Leftrightarrow (p_j, q_j)\neq 1.$$

In case R_j is finite, it is a singleton set, i.e., $R_j=\{a^{r_j}\}$. Let s_j be a divisor of r_j greater than one and smaller than itself. Then putting

$$F=\{(p_j, q_j);\ R_j \text{ is infinite}\}\cup\{s_j;\ R_j \text{ is finite}\},$$

we have the converse.

EXAMPLE 2.5. *$\{a^{k!};\ k\neq 2\}$ is regularly separable from P, but $\{a^{n^2};\ n\geqq 0\}$ not.*

3. Languages containing no infinite regular set.

In this section we show the existence of languages which are not regularly separable from any infinite language. For this purpose, first we consider the class of *rigid* languages. Here we call an infinite language *rigid* if it contains no infinite regular set.

LEMMA 3.1. *For any words u, v, w, x and y in A^* with $vx\neq\varepsilon$, a language*

$$L=\{uv^nwx^ny;\ n\geqq 0\} \tag{3.1}$$

is either regular or rigid.

PROOF. In case v or x is empty, obviously L is regular. Hence it suffices to consider the case in which neither v nor x is empty. Suppose that L is not rigid, i.e., L contains an infinite regular set R. Then R can be written as

$$R=\{uv^{n_i}wx^{n_i}y;\ i\geqq 0\}.$$

for some infinite chain $n_0<n_1<n_2<\cdots$ of integers. And from the finiteness of the number of right invariant equivalence classes induced by the regular set R, a word $uv^{n_i}wx^{n_k}y$ with $i<k$ must be in R. Since R is a subset of L, there exists an

integer j $(i<j<k)$ such that

$$uv^{n_i}wx^{n_k}y = uv^{n_j}wx^{n_j}y\,,$$

from which we have

$$wx^{n_k-n_j} = v^{n_j-n_i}w\,. \tag{3.2}$$

Here we may assume without loss of generality that the value of $k-i$ is sufficiently large or

$$l(w) < l(x^{n_k-n_j})\,, \qquad l(x^{n_j-n_i})\,.$$

Hence we have a solution of (3.2):

$$z^{n_k-n_j} = zw\,, \qquad v^{n_j-n_i} = wz\,. \tag{3.3}$$

Now noticing on that every integer n can be written as

$$n = (n_k-n_j)(n_j-n_i)p+q$$

for some p and q with $0 \leqq q < (n_k-n_i)(n_j-n_i)$, we have

$$\begin{aligned} uv^nwx^ny &= uv^qv^{(n_k-n_j)(n_j-n_i)p}wx^{(n_k-n_j)(n_j-n_i)p}x^qy \\ &= uv^q(wz)^{(n_j-n_i)p}w(zw)^{(n_k-n_j)p}x^qy \\ &= uv^q(wz)^{(n_k-n_i)p}wx^qy\,. \end{aligned}$$

Thus we have, putting $r=(n_k-n_j)(n_j-n_i)-1$,

$$L = \bigcup_{q=0}^{r} uv^q((wz)^{n_k-n_i})^*wx^qy\,,$$

which shows the regularity of L. Hence if L is not rigid, then L is regular.

Note that the language L in (3.1) is a typical context-free language with which we are familiar in the *uvwxy*-theorem.

EXAMPLE 3.1. *A language* $\{a^nb^n;\ n>0\}$ *is rigid.* This follows immediately by the above lemma, but the following example can not be proved by the lemma:

EXAMPLE 3.2. *A language* $L=\{a^nb^ma^mb^n;\ n,\ m>0\}$ *is rigid.*

PROOF. Suppose that L is not rigid. Then L must contain a regular set of the form xy^*z for some words x, y and z with $y\neq\varepsilon$. In case the word y contains both symbols a and b, then xy^*z must contain words not in L. Hence y must consist of only one symbol, but also in this case xy^*z must contain words not in L.

Now let us consider languages which are not regularly separable from any infinite language. We denote by $RS(L)$ *the class of infinite languages over a fixed alphabet which are regularly separable from the* L. For our purpose it suffices to see the emptiness of $RS(L)$. As a trivial example for which $RS(L)$ is empty, we may choose $L=A^*-F$, where F is a finite subset of A^*.

THEOREM 3.1.

(1) *There exists nonregular context-free language* L *for which* $RS(L)$ *is empty.*

(2) *For any context-free language* L *over an alphabet* A

$$RS(L)\cup RS(L^c) \neq \phi\,, \tag{3.4}$$

where L^c *is the complement of* L *with respect to* A^*.

PROOF. (1) A language $L=\{a^nb^n;\ n>0\}^c$ is nonregular context-free. And L^c is rigid as shown in the example 3.1. Hence $RS(L)$ is empty.

(2) Let L be a rigid context-free language over an alphabet A with at least two elements. Then language $L\cap a^*$ (a in A) is finite, for otherwise it becomes an infinite regular set, which is impossible. Hence there exists an integer r for which an infinite regular set a^ra^* is contained in L^c, which is again impossible. Thus if $RS(L)$ is empty then $RS(L^c)$ is not empty.

THEOREM 3.2. *There exists a context-sensitive language L such that*

$$RS(L)\cup RS(L^c)=\phi\,. \tag{3.5}$$

PROOF. Let an enumeration of all elements in $A^+\times A^+\times A^+$ be

$$\{(x_i, y_i, z_i);\ x_i, y_i, z_i \text{ in } A^+ \ \&\ i>0\}\,, \tag{3.6}$$

and let $k_1=1$ and for each $i(\geqq 2)$

$$k_i=\min\{r;\ l(x_iy_i^rz_i)>l(x_{i-1}y_{i-1}^{k_{i-1}+1}z_{i-1})\}\,.$$

Then the language

$$L=\{x_iy_i^{k_i}z_i;\ i\geqq 1\} \tag{3.7}$$

satisfies the relation (3.5). Suppose that L contains an infinite regular set. Then L contains a regular set xy^*z for some non-empty words x, y and z. Since this triplet (x, y, z) must be in the list (3.6), there must be an integer i such that $(x, y, z)=(x_i, y_i, z_i)$. However, by the definition of L, $x_iy_i^{k_i}z_i$ is in L but $x_iy_i^{k_i+1}z_i$ not. Hence xy^*z is not contained in L. By this contradiction, L can not contain any infinite regular set. Similarly L^c can not also contain any infinite regular set. Thus the relation (3.5) holds for the L.

Now we present concretely an enumeration (3.6) and show that L in (3.7) under the enumeration is recognizable by a deterministic linear bounded automaton. To do so we use some notations:

Let $A=\{a_1, a_2, \cdots, a_p\}$ be an alphabet, and ν be a bijection from the set of integers from 1 to p onto A defined by $\nu(i)=a_i$ for each i $(1\leqq i\leqq p)$. Noticing that every positive integer n can be uniquely represented as

$$n=i_m\cdot p^m+i_{m-1}\cdot p^{m-1}+\cdots+i_1\cdot p+i_0 \tag{3.8}$$

for some $i_0, i_1, \cdots, i_m$ with $1\leqq i_j\leqq p$, we can extend ν by

$$\nu(n)=\nu(i_m)\nu(i_{m-1})\cdots\nu(i_1)\nu(i_0)\,. \tag{3.9}$$

This extended mapping ν is now a bijection from the set of all positive integers N onto A^+. By μ we denote the inverse function of ν. And let τ be a bijection from $N\times N$ onto N defined by

$$\tau(m, n)=\frac{1}{2}(m^2+2mn+n^2-m-3n)+1\,. \tag{3.10}$$

We extend τ from $N\times N\times N$ onto N by

$$\tau(r, s, t)=\tau(\tau(r, s), t)\,. \tag{3.11}$$

By τ_1, τ_2 and τ_3 we denote functions of one variable which yield the inverse mappings to τ, i. e.,

$$\tau(\pi_1(n),\ \pi_2(n),\ \pi_3(n)) = n \tag{3.12}$$

for all n.

By using these functions, we can enumerate all the elements of $A^+ \times A^+ \times A^+$ as

$$\{(\nu(\pi_1(n)),\ \nu(\pi_2(n)),\ \nu(\pi_3(n)))\ ;\ n \geqq 1\}\ . \tag{3.13}$$

Here we should notice on that

$$l(\nu(\pi_i(\mu(w)))) < l(w) \tag{3.14}$$

for all w in A^+ and $i = 1, 2, 3$.

Finally let us construct a linear bounded automaton M to recognize the L in (3.7) under the enumeration (3.13). The tape of M is one divided into seven tracks as shown in the Fig. 1. The track T is to keep the original input word and the

T	*input* w
Tc	$\nu(i)$
Tx	x_i
Ty	y_i
Tz	z_i
Tr	y_i^r
To	$x_{i-1}y_{i-1}^{k_{i-1}}z_{i-1}$

Fig. 1. Tape of M.

other six tracks are for the enumeration. The behaviour of M can be described as in the Fig. 2. Now, noticing on the relation (3.14) and that the composite functions $\nu\circ\mu$, $\nu\circ\pi_1\circ\mu$, $\nu\circ\pi_2\circ\mu$ and $\nu\circ\pi_3\circ\mu$ are computable within the given tapes (i. e., arguments), it is easy to see that every operation or test in the flow diagram can be performed by a deterministic linear bounded automaton, and hence the whole diagram can be also done by a deterministic linear bounded automaton (Consult [2] for construction). And we can easily verify that M recognizes L, i. e., $L(M) = L$.

These two theorems can be restated, from the viewpoint of covering of A^*, as follows:

COROLLARY 3.1. *A^* can not be covered by any finite number of rigid context free languages.*

COROLLARY 3.2. *A^* can be covered by two rigid context-sensitive languages.*

The rest of this section is devoted to give a fundamental properties of rigid context-free languages.

THEOREM 3.3. *If L and L' are rigid context-free languages, then* (1) *$L \cap L'$,* (2) *$L \cup L'$ and* (3) *$L \cdot L'$ are ridid, but* (4) *L^* and L^c are not.*

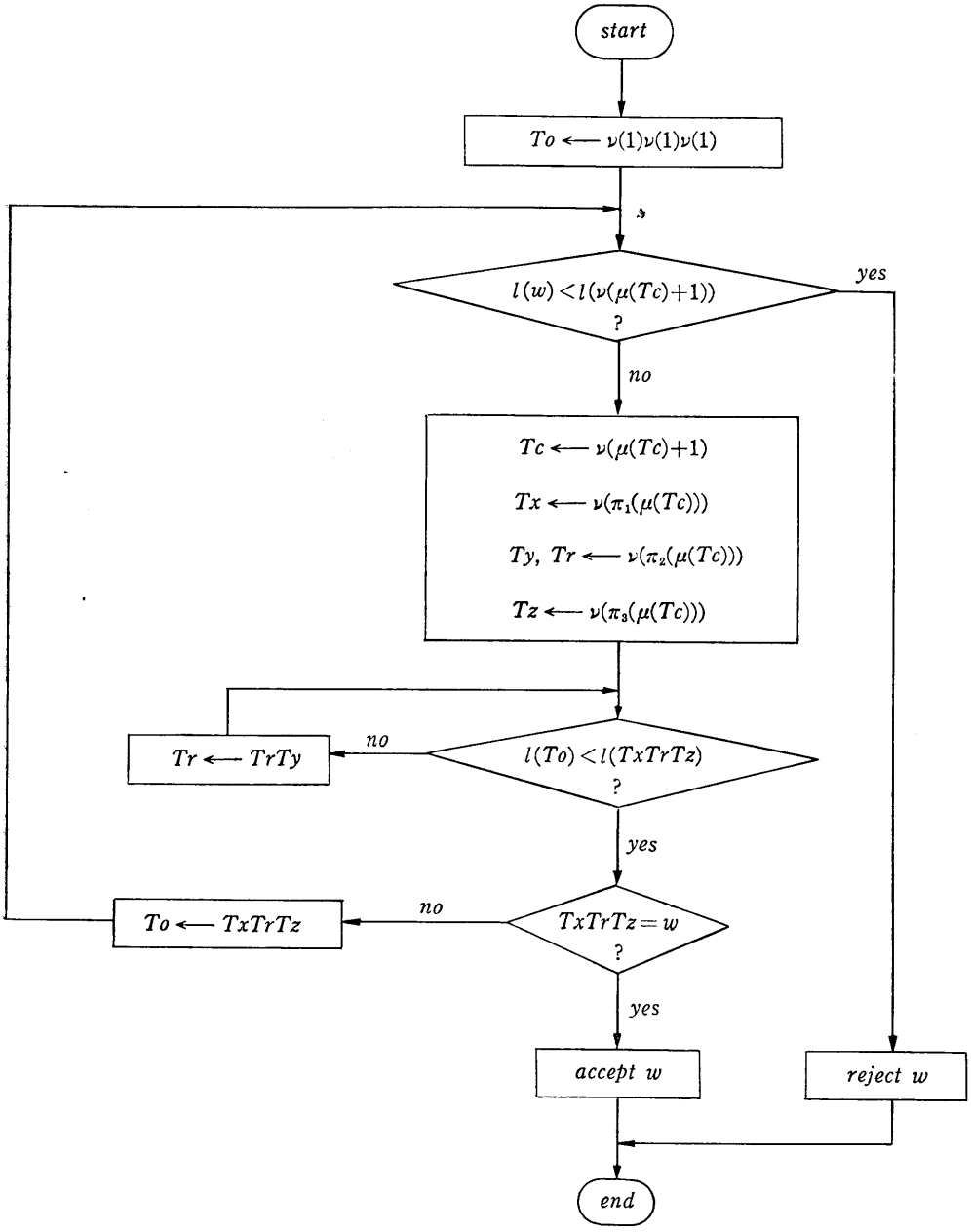

Fig. 2. Behaviour of M with an input word w.

PROOF. (1) is trivial. (2) Suppose that $L \cup L'$ is not rigid. Then it contains a regular set xy^*z for some words x, y, z with $y \neq \varepsilon$. And then at least one of the two languages $L \cap xy^*z$ and $L' \cap xy^*z$ are infinite regular, by the fact that

(3.15) $$xy^*z = (L \cap xy^*z) \cup (L' \cap xy^*z)$$

and by Theorem 4.1 in the next section. Hence L or L' is not rigid.

(3) Suppose that $L \cdot L'$ is not rigid. Then it contains a regular set xy^*z with $y \neq \varepsilon$. Consider a language

$$\bar{L} = \{u\,;\ (\exists v)(uv \text{ in } xy^*z \ \&\ u \text{ in } L \ \&\ v \text{ in } L')\}\,, \tag{3.16}$$

which is a subset of L. In case $\bar{L}$ is finite, there exist a positive integer n and a word u in $\bar{L}$ such that

$$\partial_u xy^*y^n z \subset L'\,, \tag{3.17}$$

the lefthand set of which is infinite regular.

Now let us consider the case in which $\bar{L}$ is infinite. Let $A' = \{a'\,;\ a \text{ in } A\}$ be an alphabet disjoint from A, g be a substitution defined by $g(a) = \{a, a'\}$ for every a in A, and h be a homomorphism defined by $h(a) = \varepsilon$ and $h(a') = a$ for every a and a'. Then we have

$$\bar{L} = h(g(L) \cdot L' \cap g(xy^*z))\,, \tag{3.18}$$

which shows that $\bar{L}$ is context-free, since L and L' are context-free, xy^*z is regular, g is a substitution by a finite set and h is a homomorphism. From the infiniteness of $\bar{L}$, there exists an initial subword y' of y for which $\bar{L} \cap xy^*y'$ is infinite. Since $\bar{L}$ is contex-free, the language $\bar{L} \cap xy^*y'$ is regular by Theorem 4.1. Thus we have an infinite regular set contained in L, which contradicts the assumption.

(4) For a word w in L, w^* is contained in L^*. (5) By the proof of Theorem 3.1 (1).

4. Regular sets containing no nonregular context-free language.

It is well-known that every context-free language over an alphabet with a single element is regular. In this section we extend this result to obtain a necessary and sufficient condition of regularity of context-free languages, and then prove the solvability of a decision problem questioning whether a given regular set contains nonregular context-free languages.

First we start with pointing out an elementary property.

LEMMA 4.1. *Let L, L_1 and L_2 be regular sets.*

(1) *$L_1 \cup L_2$ contains no nonregular context-free language if and only if so do both L_1 and L_2.*

(2) *If $L_1 \cdot L_2$ contains no nonregular context-free language then so do both L_1 and L_2.*

(3) *If L^* contains no nonregular context-free language, then so does L.*

PROOF. Immediate.

Note that converses of (2) and (3) are not always true. In fact, putting $L = \{a, b\}$, $L_1 = a^*$ and $L_2 = b^*$, both languages $L_1 \cdot L_2$ and L^* contain a nonregular context-free language $\{a^n b^n\,;\ n > 0\}$.

Now using the well-known result, we prove a sufficient condition:

THEOREM 4.1. *For a regular set L over an alphabet A, if there exist an integer n and words x_i, y_i, z_i in A^* with $1 \leqq i \leqq n$ such that*

(4.1) $$L \subseteqq \bigcup_{i=1}^{n} x_i y_i^* z_i ,$$

then L contains no nonregular context-free language.

PROOF. First we prove the theorem for $n=1$. Let $\bar{L}$ be a context-free language contained in L. Then a language

$$L' = \{w;\ x_1 w z_1 \text{ in } \bar{L}\}$$

is context-free, hence

$$L'' = \{a^{l(w)};\ w \text{ in } L'\}$$

is regular, where a is a symbol. Let b and c be two distinct symbols from a. Then $bL''c$ is also regular. Now let

$$y_1 = a_1 a_2 \cdots a_p \qquad (a_i \text{ in } A),$$

and

$$S = (K, \{a, b, c\}, A, \delta, \lambda, s_0)$$

be a generalized sequential machine defined by

$$K = \{s_0, s_1, \cdots, s_{p+2}\},$$

$$\delta(s_0, b) = s_1$$

$$\delta(s_i, a) = s_{i+1} \qquad (i = 1, 2, \cdots, p-1)$$

$$\delta(s_p, a) = s_1$$

$$\delta(s_p, c) = s_{p+1}$$

$$\delta(s, d) = s_{p+2} \qquad \text{for any other pair } (s, d);$$

$$\lambda(s_0, b) = x$$

$$\lambda(s_i, a) = y \qquad (i = 1, 2, \cdots, p)$$

$$\lambda(s_p, c) = z$$

$$\lambda(s, d) = \varepsilon \qquad \text{for any other pair } (s, d).$$

Then we have easily $\bar{L} = S(bL''c)$, which shows the regularity of $\bar{L}$.

Now suppose (4.1) and that $\bar{L}$ is a context-free language contained in L. Then

$$\bar{L} = \bigcup_{i=1}^{n} (\bar{L} \cap x_i y_i^* z_i),$$

and, for every i, language $\bar{L} \cap x_i y_i^* z_i$ is regular as we have just proved. Hence so is $\bar{L}$ itself.

In order to prove the converse of the theorem we prepare some lammas.

LEMMA 4.2. *Let a language R contain no nonregular context-free language. Then so does R^* if and only if R is commutative (i.e., $uv = vu$ for any words u, v in R).*

PROOF. Suppose that R is commutative. Then there exists a word w for which $R \subseteqq w^*$ (See p. 169 in [3]). Hence $R^* \subseteqq w^*$, which proves the sufficiency by Theorem 4.1. Suppose, conversely, that R is not commutative. Then there exist two words u, v in R such that $uv \neq vu$. By using this pair of words, we have a

context-free language contained in R^*

$$L = \{(uv)^n(vu)^n\,;\ n > 0\}\ ,$$

which is not regular, for we can construct a generalized sequential machine which maps L onto a nonregular context-free language $\{a^nb^n\,;\ n > 0\}$, where a, b are symbols.

COROLLARY 4.1. *Let a language R contain no nonregular context-free language. Then so does R^* if and only if*

$$R^* \subseteqq \bigcup_{i=1}^{n} x_i y_i^* z_i \tag{4.2}$$

for some words x_i, y_i, z_i $(1 \leqq i \leqq n)$.

PROOF. It is sufficient to prove that if R is not commutative, then R^* can not be covered by the righthand side of (4.2) for any choice of words x_i, y_i, z_i. This follows, however, immediately from the nonregularity of L in the above proof and Theorem 4.1.

LEMMA 4.3. *Let u, v, w, x and y be words. Then a language ux^*vy^*w contains no nonregular context-free langnage if and only if there exist integers m, n, p and q $(m \neq p)$ such that*

$$ux^mvy^nw = ux^pvy^qw\ . \tag{4.3}$$

PROOF. Suppose that the equation (4.3) holds. In case $m > p$ and $n \leqq q$, we have

$$x^{m-p}v = vy^{q-n}\ . \tag{4.4}$$

By repeatedly using the relation (4.4), each word in ux^*vy^*w can be written in a form

$$\begin{aligned} ux^kvy^hw &= ux^{(m-p)r+t}vy^hw \qquad (0 \leqq t < m-p,\ r \geqq 0) \\ &= ux^tvy^{h+r(q-n)}w\ . \end{aligned}$$

Hence we have

$$ux^*vy^*w = \bigcup_{t=0}^{m-p-1} ux^tvy^*w\ . \tag{4.5}$$

And by Theorem 4.1, ux^*vy^*w can not contain nonregular context-free language.

Conversely suppose that (4.3) does not hold. And suppose that a language

$$L = \{ux^nvy^nw\,;\ n > 0\} \tag{4.6}$$

is regular. Then from the finiteness of the number of right invariant equivalence classes induced by the L, there exists a class in which word ux^m and ux^n $(m \neq n)$ are contained, and two words ux^mvy^mw and ux^nvy^mw must be in one class, i. e., these two words must be contained in the L, which is impossible, because of our first assumption. Hence L is not regular, or the language ux^*vy^*w contains nonregular context-free language.

COROLLARY 4.2. *The language ux^*vy^*w does not contain nonregular context-free language if and only if*

$$ux^*vy^*w \subseteqq \bigcup_{i=1}^{n} x_i y_i^* z_i \tag{4.7}$$

for some words x_i, y_i, z_i.

PROOF. It is sufficient to see that the language L in the proof of Lemma 4.3 can not be covered by any regular set of the form of the righthand side of (4.7). But this follows immediately from Theorem 4.1.

THEOREM 4.2. *If a regular set* L *over* A *does not contain any nonregular context-free language, then there exist an integer* n *and words* x_i, y_i, z_i *in* A^* *with* $1 \leqq i \leqq n$ *such that*

$$L \subseteqq \bigcup_{i=1}^{n} x_i y_i^* z_i .$$

PROOF. A proof is done easily by an induction on the number of * operations defining the regular set L, using Lemma 4.1 and the last corollaries.

Note that, by the proofs of the lemmas, if a regular set contains a nonregular context-free language, then as such a language we can choose one in a form $\{ux^n vy^n w\,;\ n > 0\}$.

Finally we give a solvable decision problem.

LEMMA 4.4. *Let* R *be a regular set. Then it is solvable to determine whether* R^* *contains nonregular context-free languages.*

PROOF. By the above lemmas, we have immediately:

R^* *contains no nonregular context-free language*

$\Leftrightarrow$ R *is commutative*

$\Leftrightarrow$ $R \subseteqq w^*$ *for some word* w

$\Leftrightarrow$ $R^* \subseteqq w^*$ *for some word* w

$\Leftrightarrow$ $R^* \subseteqq w^*$ *for the shortest nonempty word* w *in* R.

LEMMA 4.5. *For a regular set of the form* ux^*vy^*w, *it is solvable to determine whether it contains nonregular contex-free languager.*

PROOF. By the proof of Lemma 4.3, we have an equivalence:

ux^*vy^*w *contains no nonregular set*

$\Leftrightarrow$ *there exist integers* m *and* n *such that* $x^m v = vy^n$.

And we have:

$$x^m v = vy^n \Leftrightarrow x^{l(y)} v = vy^{l(x)} . \tag{4.8}$$

It is sufficient to see this from left to right. Noticing on the fact that if $x^m v = vy^n$ then $x^{pm} v = vy^{pn}$ for any integer p, we may assume in (4.8) that

$$m = k \cdot l(y) , \qquad n = k \cdot l(x) .$$

In case $l(v) > l(x) \cdot l(y)$, by using subwords x_1 and x_2 of $x^{l(y)}$ such that $v = x^{l(y)\cdot r} x_1$ and $x^{l(y)} = x_1 x_2$, and noticing on that $y^{l(x)} = x_2 x_1$, we have

$$\begin{aligned} x^{l(y)} v &= x^{l(y)} x^{l(y)\cdot r} x_1 \\ &= x^{l(y)\cdot r} x_1 x_2 x_1 \\ &= vy^{l(x)} . \end{aligned} \tag{4.9}$$

In case $l(v) \leqq l(x) \cdot l(y)$, by using a subword x' such that $x^{l(y)} = vx'$, we have $y^{l(y)} = x'v$ and

$$x^{l(y)}v = vx'v = vy^{l(x)}. \tag{4.10}$$

Hence we have the implication of (4.8) from left to right.

THEOREM 4.3. *For a given regular set R, it is solvable to determine whether R contains nonregular context-free languages.*

PROOF. Proof is done easily by an induction on the number of regularity operations, using the last two lemmas.

This theorem is an interesting contrast with the well-known result on context-free languages:

THEOREM (*See p.* 132 *in* [3]). *For a given context-free language L, it is unsolvable to determine whether L contains infinite regular sets.*

References

[1] K. KOBAYASHI, *Structural complexity of context-free languages*, Information and Control, **18** (1971), 299-310.

[2] S. ARIKAWA, *On the length functions of languages recognizable by linear bounded automata*, Mem. Fac. Sci., Kyushu Univ., Ser. A, **23** (1969), 12-27.

[3] S. GINSBURG, The mathematical theory of context-free languages, McGraw Hill, New York (1966).