Theoretical Computer Science 2 (1976) 147–154.

REGULARITY-PRESERVING RELATIONS*

J. I. SEIFERAS

The Pennsylvania State University, University Park, Pa. 16802, USA

R. McNAUGHTON

Rensselaer Polytechnic Institute, Troy, N. Y. 12181, USA

Communicated by A. Meyer
Received January 1975
Revised May 1975

1. Introduction

In 1962, Ginsburg raised the question of whether the regularity of the language L implies the regularity of

$$\text{FIRST-HALF}(L) = \{x \mid \exists y \ (|y| = |x| \ \& \ xy \in L)\},$$

where $|w|$ denotes the length of the word w. In response, Yamada [8] and Chang independently proved that it does; and Stearns and Hartmanis [7] subsequently published a study of such "proportional removals" from regular languages. Kosaraju [2] and Seiferas [5] later extended these results to certain less-than-proportional "removals" from regular languages. In this note, we finally give a complete characterization of which prefix removals of regular languages are regular.

Definition. For each binary relation r on the set N of nonnegative integers and each language L, define

$$P(r, L) = \{x \mid \exists y (r(|x|, |y|) \ \& \ xy \in L\},$$

where $r(i, j)$ indicates that the ordered pair (i, j) is in the relation r. We say that a relation r is *regularity-preserving* if $P(r, L)$ is regular for every regular language L. By identifying each function $f : \mathrm{N} \to \mathrm{N}$ with the relation $\{(n, f(n)) \mid n \in \mathrm{N}\}$, we give meaning also to $P(f, L)$ and to *regularity-preserving functions.*

Examples. The function f defined by $f(n) = n$ is regularity-preserving because $P(f, L) = \text{FIRST-HALF}(L)$.

For $f(n) = n^2 - n$, we get

* This work was supported in part by the National Science Foundation under Research Grants GJ-34671 (first author) and GJ-35293 (second author).

$$P(f, L) = \{x \mid \exists y\ (|xy| = |x|^2 \ \&\ xy \in L)\},$$

which we might refer to as SQUARE-ROOT(L) . Note that the question of whether SQUARE-ROOT(L) is regular for every regular language L hinges on whether $n^2 - n$, roughly the *inverse* of $n^{\frac{1}{2}}$, is regularity-preserving.

Similarly, we might define

$$\text{ROUNDED-SQUARE-ROOT}(L) = P(\{(\lfloor n^{\frac{1}{2}} \rfloor,\ n - \lfloor n^{\frac{1}{2}} \rfloor)\}, L),$$

where $\lfloor x \rfloor$ denotes the greatest integer not exceeding x.

Definition. A set A of nonnegative integers is *u.p.* (*ultimately periodic*) if there is a positive integer p such that the following holds for all but finitely many $n \in \mathbf{N}$: $n \in A \Leftrightarrow n + p \in A$. A relation r on the nonnegative integers is *u.p.-preserving* if $r^{-1}(A) = \{i \mid (\exists j \in A)\ r(i, j)\}$ is u.p. for every u.p. set A.

Our main result is that a relation is regularity-preserving if and only if it is u.p.-preserving. Furthermore, we show that the u.p.-preserving relations include a class of functions (the "essentially increasing u.p. reducible" functions defined below) shown by Siefkes [6] to be very rich. For each integer $k > 1$, for example, the functions n^k and k^n are in the class; and the sums, products, compositions, and certain iterations (e.g., the exponential stack of 2's of height n) of functions in the class are in the class.

Remark. Kosaraju [4] has recently characterized the class of functions f having the property that $P(f, L)$ is *context-free* for every *context-free* language L. These "c.f.-preserving" functions turn out to be precisely the u.p.-preserving functions f that satisfy the additional condition that the following set is finite for each k: $\{f(n) \mid f(n) \leq kn\}$.

2. Characterization as u.p.-preserving relations

We assume the reader is familiar with the definition and basic properties of regular languages. (These can be found, along with further references, in [1, Chapter 3].) We recall two well-known results about the relationship between regular languages and u.p. sets.

Lemma 1. *If L is regular, then $\{|x| \mid x \in L\}$ is u.p.*

Lemma 2. *If A is u.p., then $\{x \in \Sigma^* \mid |x| \in A\}$ is regular, for each finite alphabet Σ.*

Theorem 1. *A relation is regularity-preserving if and only if it is u.p.-preserving.*

Proof. (only if) Assume r is regularity-preserving. Suppose A is u.p. By Lemma 2, $L = \{1^n \mid n \in A\}$ is regular. Therefore, $L' = 0^*1 \cap P(r, 0^*1L)$ is regular. By Lemma 1, $\{|x| \mid x \in L'\}$ is u.p. But $\{|x| \mid x \in L'\} = r^{-1}(A) - \{0\}$; so $r^{-1}(A)$ is u.p., too.

(if) Assume r is u.p.-preserving. Suppose $L \subseteq \Sigma^*$ is regular. Recall that in the proof of Nerode's theorem [1, Theorem 3.1] we take a deterministic finite automaton M that accepts L and partition Σ^* into the equivalence classes of the equivalence relation "leads to the same state of M as". This gives a partition of Σ^* into finitely many regular sets $L_1, \cdots, L_k$ (one for each accessible state of M) such that $\{y \mid x_1 y \in L\} = \{y \mid x_2 y \in L\}$ whenever x_1, x_2 lie in the same block L_i. For each i define R_i to be the regular set $\{y \mid xy \in L\}$ obtained for every $x \in L_i$. Because $\Sigma^* = L_1 \cup \cdots \cup L_k$, we have

$$P(r, L) = (P(r, L) \cap L_1) \cup \cdots \cup (P(r, L) \cap L_k).$$

Since r is u.p.-preserving, the regularity of

$$\begin{aligned} P(r, L) \cap L_i &= \{x \in L_i \mid \exists y\ (r(|x|, |y|)\ \&\ xy \in L)\} \\ &= \{x \in L_i \mid (\exists y \in R_i)\ r(|x|, |y|)\} \\ &= \{x \in \Sigma^* \mid (\exists y \in R_i)\ r(|x|, |y|)\} \cap L_i \\ &= \{x \in \Sigma^* \mid |x| \in r^{-1}(\{|y| \mid y \in R_i\})\} \cap L_i \end{aligned}$$

for each i follows by Lemmas 1, 2. □

3. U.p.-degenerating relations

Consider the relation $r = \{(\lfloor n^{\frac{1}{2}} \rfloor, n - \lfloor n^{\frac{1}{2}} \rfloor) \mid n \in \mathbb{N}\}$. If r is u.p.-preserving, then the regularity of L implies the regularity of ROUNDED-SQUARE-ROOT(L), by Theorem 1. In fact r is quite trivially u.p.-preserving:

If A is u.p., then

$$(1) \qquad r^{-1}(A) \text{ is } \begin{cases} \text{finite} & \text{if } A \text{ is finite,} \\ \text{cofinite} & \text{if } A \text{ is infinite.} \end{cases}$$

[For each i, note that $r(i, j)$ holds for every j with $i^2 - i \leq j \leq i^2 + i$. Thus if A is infinite and u.p. with $(\forall n \geq n_0)\ (n \in A \Leftrightarrow n + p \in A)$, then $(\exists j \in A) r(i, j)$ holds for every i so large that $i^2 - i \geq n_0$, $2i \geq p$.] Let us call any relation satisfying (1) *u.p.-degenerating.*

Theorem 2. *If r_1, r_2 are u.p.-degenerating relations, then $P(r_1, L) - P(r_2, L)$ is finite for every regular language L.*

Proof. Let L be any regular language over Σ. In proving Theorem 1, we found regular sets $L_1, \cdots, L_k$ and $R_1, \cdots, R_k$ over Σ such that, for *any* relation r on $\mathbb{N}$,

$$P(r, L) = \bigcup_{i=1}^{k} (\{x \in \Sigma^* \mid |x| \in r^{-1}(\{|y| \mid y \in R_i\})\} \cap L_i).$$

For r_1, r_2 u.p.-degenerating, define

$$L_{i,j} = \{x \in \Sigma^* \mid |x| \in r_j^{-1}(\{|y| \mid y \in R_i\})\} \qquad (1 \leq i \leq k, 1 \leq j \leq 2).$$

By definition (and Lemmas 1, 2),

$$L_{i,j} \text{ is } \begin{cases} \text{finite} & \text{if } R_i \text{ is finite,} \\ \text{cofinite (in } \Sigma^*) & \text{if } R_i \text{ is infinite;} \end{cases}$$

so $L_{i,1} - L_{i,2}$ is finite for each i. But

$$\begin{aligned} P(r_1, L) - P(r_2, L) &= \bigcup_{i=1}^{k} (L_{i,1} \cap L_i) - \bigcup_{i=1}^{k} (L_{i,2} \cap L_i) \\ &\subseteq \bigcup_{i=1}^{k} ((L_{i,1} - L_{i,2}) \cap L_i) \\ &\subseteq \bigcup_{i=1}^{k} (L_{i,1} - L_{i,2}). \ \square \end{aligned}$$

Remark. Consider any fixed regular language L. By Nerode's theorem again, there are only finitely many distinct finite sets of the form $\{y \mid xy \in L\}$. If we take e to exceed the lengths of all strings in these sets and define a u.p.-degenerating relation

$$r_L = \{(i, j) \mid j - i \geq e\},$$

then we get

$$P(r_L, L) = \{x \mid \{y \mid xy \in L\} \text{ is infinite}\}.$$

By Theorem 2, therefore, every language in $\{P(r, L) \mid r \text{ is u.p.-degenerating}\}$ is merely a finite variation of the regular language $\{x \mid \{y \mid xy \in L\} \text{ is infinite}\}$.

Examples. Define

$$r_1 = \{(\lfloor n^{\frac{1}{3}} \rfloor, n - \lfloor n^{\frac{1}{3}} \rfloor)\}, \qquad \text{ROUNDED-CUBE-ROOT}(L) = P(r_1, L);$$

$$r_2 = \{(\lfloor \log_2 n \rfloor, n - \lfloor \log_2 n \rfloor)\}, \quad \text{ROUNDED-BASE-2-LOG}(L) = P(r_2, L);$$

$$r_3 = \{(\log^* n, n - \log^* n)\}, \qquad \text{LOG-STAR}(L) = P(r_3, L),$$

where

$$\log^* n = \min \Big\{ k \mid \underbrace{2^{2^{\cdot^{\cdot^{2}}}}}_{k} \geq n \Big\}.$$

Like $\{(\lfloor n^{\frac{1}{2}} \rfloor, n - \lfloor n^{\frac{1}{2}} \rfloor)\}$, the relations r_1, r_2, r_3 are u.p.-degenerating. If L is regular, therefore, then ROUNDED-SQUARE-ROOT(L) is regular, and the following holds for all sufficiently long strings x:

$$\begin{aligned} x \in \text{ROUNDED-SQUARE-ROOT}(L) &\Leftrightarrow x \in \text{ROUNDED-CUBE-ROOT}(L) \\ &\Leftrightarrow x \in \text{ROUNDED-BASE-2-LOG}(L) \\ &\Leftrightarrow x \in \text{LOG-STAR}(L) \\ &\Leftrightarrow \{y \mid xy \in L\} \quad \text{is infinite.} \end{aligned}$$

4. The class of regularity-preserving functions

By Theorem 1, the regularity-preserving *functions* are precisely the u.p.-preserving ones. In an entirely different context, Siefkes [6] happens to have studied the closure properties of classes of functions closely related to the class of u.p.-preserving functions.

Definition (Siefkes [6]). A function $f : \mathbf{N} \to \mathbf{N}$ is *u.p. reducible* if, for every modulus m, there is a period p such that the following holds for all but finitely many $n \in \mathbf{N}$: $f(n) \equiv f(n+p) \pmod m$ (i.e., $f(n) - f(n+p)$ is divisible by m). A function $f : \mathbf{N} \to \mathbf{N}$ is *essentially increasing* if, for every k, $f(n) \geq k$ for all but finitely many $n \in \mathbf{N}$.

Let

$$F = \{f \mid f \text{ is essentially increasing and u.p. reducible}\},$$

$$G = \{f \mid f \text{ is u.p.-preserving}\},$$

$$H = \{f \mid f \text{ is u.p. reducible}\}.$$

Arguments in [6] show that F and H are very rich classes, and the results of this section show that G lies properly between them (i.e., $F \subsetneqq G \subsetneqq H$). The particular result $F \subsetneqq G$ shows that G is richer than F, so that Theorem 1 is stronger than the earlier result of [5].

Theorem 3. $F \subsetneqq G$.

Proof. ($\subseteq$) Assume f is essentially increasing and u.p. reducible. Suppose A is u.p. Take n_0, m such that $(\forall n \geq n_0)\ (n \in A \Leftrightarrow n + m \in A)$. Thus, $(\forall n_1, n_2 \geq n_0)$ $(n_1 \equiv n_2 \pmod m \Rightarrow (n_1 \in A \Leftrightarrow n_2 \in A))$. Take n_0', p such that $(\forall n \geq n_0')\ (f(n) \geq n_0$ & $f(n) \equiv f(n+p) \pmod m)$. Then

$$\begin{aligned} n \geq n_0' &\Rightarrow f(n), f(n+p) \geq n_0 \;\&\; f(n) \equiv f(n+p) \pmod m \\ &\Rightarrow (f(n) \in A \Leftrightarrow f(n+p) \in A) \\ &\Rightarrow (n \in f^{-1}(A) \Leftrightarrow n + p \in f^{-1}(A)); \end{aligned}$$

i.e., $f^{-1}(A)$ is u.p.

($\neq$) Define

$$f(n) = \begin{cases} 0 & \text{if } n \text{ is even,} \\ n! & \text{if } n \text{ is odd.} \end{cases}$$

Clearly $f \notin F$. For A u.p., it is easy to see that $f^{-1}(A)$ contains either all or no even numbers and either only finitely many or all but finitely many odd numbers; in either case, $f^{-1}(A)$ is u.p., so $f \in G$. □

Examples. The functions

$$f_1(n) = n^2, \quad f_2(n) = n^3, \quad f_3(n) = 2^n, \quad f_4(n) = \underbrace{2^{2^{\cdot^{\cdot^{2}}}}}_{n},$$

and n are u.p. reducible; and the difference of u.p. reducible functions (if it does not go negative) is easily seen to be u.p. reducible. Thus, the essentially increasing functions $f_1(n)-n$, $f_2(n)-n$, $f_3(n)-n$, and $f_4(n)-n$ are u.p. reducible. By Theorems 3, 1, therefore, the following languages are regular if L is regular:

$$\text{SQUARE-ROOT}(L)=P(f_1(n)-n, L),$$
$$\text{CUBE-ROOT}(L)=P(f_2(n)-n, L),$$
$$\text{BASE-2-LOG}(L)=P(f_3(n)-n, L),$$
$$\text{BASE-2-HEIGHT}(L)=P(f_4(n)-n, L).$$

Theorem 4. $G \subsetneq H$.

Proof. ($\subseteq$) Assume f is u.p.-preserving. Let m be any positive integer. Then $f^{-1}(\{j \mid j \equiv n \pmod{m}\})$ is u.p. for each n, $0 \leq n \leq m-1$. If p is a common period for these m u.p. sets, then $f(n) \equiv f(n+p) \pmod{m}$ for all but finitely many n.

($\neq$) For any set A that is not u.p. (e.g., the set of primes), define

$$f(n)=\begin{cases}0 & \text{if } n \in A,\\ n! & \text{if } n \notin A.\end{cases}$$

For every m and every $n \geq m$,

$$f(n) \equiv f(n+1) \equiv 0 \pmod{m},$$

so $f \in H$. On the other hand, the singleton $\{0\}$ is certainly u.p., but $f^{-1}(\{0\})=A$ is not u.p.; so $f \notin G$. □

If we consider all functions f that are regularity-preserving with respect to just languages over a one-letter alphabet (i.e., $P(f, L)$ is regular for every regular $L \subseteq \{1\}^*$), then we get the whole class H (and hence, by Theorem 4, more than just the regularity-preserving functions as actually defined).

Theorem 5. $H=\{f \mid P(f, L)$ *is regular for every regular* $L \subseteq \{1\}^*\}$.

Proof. Let f be any function. Define an essentially increasing function g by $g(n)=n+f(n)$. Because H is closed under sum and (when the result is nonnegative-valued) difference, the following are equivalent: $f \in H$, $g \in F$, $g \in H$.

($\supseteq$) Assume $P(f, L)$ is regular for every regular $L \subseteq \{1\}^*$. Suppose A is u.p. Then

$$\begin{aligned} g^{-1}(A) &= \{i \mid g(i) \in A\} \\ &= \{i \mid i+f(i) \in A\} \\ &= \{|x| \mid x \in P(f, \{z \in \{1\}^* \mid |z| \in A\})\} \end{aligned}$$

is u.p. by Lemmas 2, 1. Thus, $g \in G \subseteq H$ (by Theorem 4); so $f \in H$.

($\subseteq$) Assume f is u.p. reducible. Then $g \in F$, so that g is u.p.-preserving by Theorem 3.

Suppose $L \subseteq \{1\}^*$ is regular. Since g is u.p.-preserving, the regularity of

$$\begin{aligned} P(f, L) &= \{x \mid \exists y\ \ (f(|x|) = |y|\ \&\ xy \in L)\} \\ &= \{x \mid \exists y\ \ (g(|x|) = |xy|\ \&\ xy \in L)\} \\ &= \{x \in \{1\}^* \mid |x| \in g^{-1}(\{|z| \mid z \in L\})\} \end{aligned}$$

follows by Lemmas 1, 2. □

Finally, we give a characterization (suggested by Kosaraju [3]) of G as a restriction of H.

Theorem 6. $G = \{f \in H \mid f^{-1}(\{j\})$ *is u.p. for every* $j\}$.

Proof. ($\subseteq$) Assume $f \in G$. By Theorem 4, $f \in H$. Since $\{j\}$ is certainly u.p. for each j, $f^{-1}(\{j\})$ is u.p. for each j.

($\supseteq$) Assume $f \in H$ and $f^{-1}(\{j\})$ is u.p. for every j. Suppose A is u.p. Take n_0, p such that

$$(\forall n \geqslant n_0)\quad (n \in A \Leftrightarrow n + p \in A);$$

then

$$\{j \in A \mid j \geqslant n_0\} = \Bigg(\bigcup_{\substack{n \in A \\ n_0 \leqslant n < n_0 + p}} \{j \mid j \equiv n \pmod p\}\Bigg) - \{j \mid j < n_0\}.$$

Then

$$\begin{aligned} f^{-1}(A) &= f^{-1}(\{j \in A \mid j < n_0\}) \cup f^{-1}(\{j \in A \mid j \geqslant n_0\}) \\ &= f^{-1}(\{j \in A \mid j < n_0\}) \cup \Bigg(f^{-1}\Bigg(\bigcup_{\substack{n \in A \\ n_0 \leqslant n < n_0 + p}} \{j \mid j \equiv n \pmod p\}\Bigg) \\ &\qquad\qquad - f^{-1}(\{j \mid j < n_0\})\Bigg) \\ &= \Bigg(\bigcup_{\substack{j \in A \\ j < n_0}} f^{-1}(\{j\})\Bigg) \cup \Bigg(\Bigg(\bigcup_{\substack{n \in A \\ n_0 \leqslant n < n_0 + p}} \{i \mid f(i) \equiv n \pmod p\}\Bigg) - \Bigg(\bigcup_{j < n_0} f^{-1}(\{j\})\Bigg)\Bigg). \end{aligned}$$

Because f is u.p. reducible, each of the sets $\{i \mid f(i) \equiv n \pmod p\}$ is u.p. By assumption, $f^{-1}(\{j\})$ is u.p. for each j. The u.p. sets are easily seen to be closed under set union and set difference, so $f^{-1}(A)$ is u.p. □

Acknowledgment

We thank Professors Albert R. Meyer and Vaughan R. Pratt for their help with earlier versions of this paper. We also thank the referees for their suggestions and corrections.

References

[1] J. E. Hopcroft and J. D. Ullman, *Formal Languages and Their Relation to Automata* (Addison-Wesley, Reading, Mass., 1969).
[2] S. R. Kosaraju, Finite state automata with markers, in *Proc. Fourth Annual Princeton Conference on Information Sciences and Systems*, Princeton, N. J. (1970) 380.
[3] S. R. Kosaraju, Regularity preserving functions, *SIGACT News* 6 (2) (April 1974) 16–17.
[4] S. R. Kosaraju, Context-free preserving functions, *Math. Systems Theory* 9 (3) (1975) 193–197.
[5] J. I. Seiferas, A note on prefixes of regular languages, *SIGACT News* 6 (1) (January 1974) 25–29.
[6] D. Siefkes, Decidable extensions of monadic second order successor arithmetic, in: *Automatentheorie und formale Sprachen*, J. Dorr and G. Hotz (eds) (Mannheim, 1970) 441–472.
[7] R. E. Stearns and J. Hartmanis, Regularity preserving modifications of regular expressions, *Information and Control* 6 (1) (March 1963) 55–69.
[8] H. Yamada, Fractionalization of regular expressions, unpublished note (1962).