Theoretical Computer Science 21 (1982) 225–230
North-Holland Publishing Company

NOTE

ON THE SYNTACTIC TRANSFORMATION SEMIGROUP OF A LANGUAGE GENERATED BY A FINITE BIPREFIX CODE

Stuart W. MARGOLIS
Department of Mathematics, University of Vermont, Burlington, VT 05401, U.S.A.

Communicated by M. Nivat
Received December 1981
Revised April 1982

Abstract. Let P be a finite biprefix code and let $X=(Q,S)$ be the syntactic transformation semigroup (ts) of P^+. We show that if $e\in S$ is an idempotent, then the ts $X_e=(Qe,eSe)$ consists of partial one to one maps. We also show that any ts of partial one to one maps divides a ts of partial one to one maps which is the syntactic ts of a finite biprefix code.

1. Introduction

Let A be a finite set. A subset P of the free semigroup A^+ is a prefix if $P\cap PA^+=\emptyset$. A suffix is defined dually and a biprefix is a set which is both a prefix and a suffix. A prefix P is complete if $P^+\cap wA^+\neq\emptyset$ for all $w\in A^+$.

It is well known that the subsemigroup P^+ generated by a prefix P is free. In fact, P^+ satisfies the following condition: If $w\in A^+$ and $P^+w\cap P^+\neq\emptyset$ then $w\in P^+$. An important tool for studying P^+ is the syntactic semigroup $S(P^+)$. We recall that $S(P^+)$ is the quotient of A^+ by the largest congruence such that P^+ is a union of classes. This study was initiated by Schutzenberger in [10] and we refer the reader to Chapter 8 of [3] for basic results. We also recall Kleene's theorem which states that a subset L of A^+ is rational (i.e. regular) if and only if $S(L)$ is finite.

Recently there have been a number of results showing how an arbitrary finite semigroup divides a semigroup of the form $S(P^+)$ where P is a rational prefix code. Indeed Schutzenberger shows [11] that any finite semigroup is a subsemigroup of $S(P^+)$ where P is a complete rational biprefix. In [6] Pin proves that any semigroup divides $S(P^+)$ for some *finite* prefix P, a result that is refined in [7] and [5].

On the other hand, it is well known that a finite complete prefix is a biprefix if and only if $S=S(P^+)$ is nil-simple, [3]. That is, for all $s\in S$, there is an n such that s^n is in the minimal ideal of S. It is easy to show that a finite semigroup S is nil-simple if and only if eSe is a group for all idempotents $e\in S$.

In view of these results, it is reasonable to ask if every finite semigroup S divides $S(P^+)$ where P is a finite biprefix. The main result of this paper shows that this is not true by proving that if P is a finite biprefix, then $eS(P^+)e$ is a subsemigroup of an inverse semigroup for any idempotent $e \in S(P^+)$. More generally, if $X = (Q, S(P^+))$ is the syntactic transformation semigroup (ts) of P^+, then $X_e = (Qe, eS(P^+)e)$ is an injective ts. That is each transformation of X_e is partial one-one. We call such a ts, locally injective.

Let $A = \{a, b\}$. We remark that the syntactic ts of $P^+ = \{a, ba\}^+$ is locally injective, so that the converse of the above result is not true. We will show however, using the techniques of [7], that any injective ts divides the syntactic ts of a finite biprefix. For other results on injective biprefixes see [2], [8], and [9].

All undefined notions and terminology can be found in [1] or [3]. In particular, an A-automaton $\mathscr{A} = (Q, A)$ is a *partial* function $Q \times A \to Q$ where Q is a finite set. The ts of $\mathscr{A}$ is the pair $X = (Q, S)$ where S is the semigroup generated by the partial functions in A.

2. The main result

Let A be a finite set and let $P \subseteq A^+$ be a rational prefix. Let $\mathscr{A} = (Q, A)$ be the minimal automaton of P^+. We recall that there is an $i \in Q$ such that $P^+ = \{w \mid iw = i\}$. More generally if $q \in Q$, let $\mathscr{A}_q = \{w \in A^+ \mid qw = q\}$. Let $P\alpha = \{v \in A^* \mid vA^+ \cap P \neq \emptyset\}$ and let $P\omega = \{v \in A^* \mid A^+v \cap P \neq \emptyset\}$.

Lemma 1. *Let P be a finite prefix and let $\mathscr{A} = (Q, A)$ be the minimal automaton of P^+. If $v \in \mathscr{A}_q$ for some $q \in Q$, then $v = xdy$ for some $x \in P\omega$, $d \in P^*$, $y \in P\alpha$. Furthermore $yx \in P \cup \{1\}$.*

Proof. We recall that the states of $\mathscr{A}$ are the sets of the form $s^{-1}P^+ = \{w \mid sw \in P^+\}$ for $s \in P\alpha$ and that $i = P^+$. Let $q = s^{-1}P^+$. Since P is finite, there exists a prefix x of v such that $sx \in P \cup \{1\}$. Therefore, $x \in P\omega$. Let d be the longest prefix of $x^{-1}v$ such that $d \in P^*$. Then $v = xdy$ for some $y \in P\alpha$. Furthermore, $qx = i$ and $iy = q$ and it follows that $yx \in P \cup \{1\}$. $\square$

Proposition 2. *Let P be a finite biprefix and let $\mathscr{A} = (Q, A)$ be the minimal automaton of P^+. Suppose there are $q, q' \in Q$, and $v \in \mathscr{A}_q \cap \mathscr{A}_{q'}$. If there is $w \in A^*$ such that $qw = q'w \neq \emptyset$, then $q = q'$.*

Proof. Since $\mathscr{A}$ is transitive, we can assume that $qw = q'w = i$ the state stabilized by P^+. By Lemma 1, v factors

$$v = xdy = x'd'y' \tag{1}$$

where

$$x, x' \in P\omega, \quad d, d' \in P^*, \quad y, y' \in P\alpha, \quad yx, y'x' \in P \cup \{1\}.$$

It follows that $iy = q$, $iy' = q'$. By our assumption on w, we have $yw \in P^*$ and $y'w \in P^*$. Without loss of generality, there is $z \in A^*$ such that $y' = zy$ by (1).

Therefore $y'x'd'y'w \in P^*$. But,

$$y'x'd'y'w = y'x'd'zyw.$$

Since $y'x'd' \in P^*$, it follows that $zyw \in P^*$ since P is a prefix. Using the fact that P is a suffix and $yw \in P^*$, we have $z \in P^*$.

Thus $q' = iy' = izy = iy = q$. $\square$

Let $X = (Q, S)$ be a ts. If $e \in S$ is an idempotent, let $X_e = (Qe, eSe)$. X is *injective*, if each $s \in S$ is partial one–one. X is *locally injective* if X_e is injective for all idempotents $e \in S$.

Theorem 3. *Let P be a finite biprefix and let $X = (Q, S)$ be the syntactic ts of P^+. Then X is locally injective.*

Proof. Let $\mathcal{A} = (Q, A)$ be the minimal automaton of P^+. Then X is the ts of $\mathcal{A}$ and $S = S(P^+)$ is the syntactic semigroup of P^+. Let $\eta : A^+ \to S$ be the syntactic morphism. Let $e = e^2 \in S$ and let $v \in e\eta^{-1}$. Assume that there are q, $q' \in Qe$ and $s \in eSe$ such that $qs = q's \neq \emptyset$. Let $w \in s\eta^{-1}$. Then $qv = q$ and $q'v = q'$ since $\{q, q'\} \subseteq Qe$. Therefore $v \in \mathcal{A}_q \cap \mathcal{A}_{q'}$ and since $qw = q'w \neq \emptyset$ Proposition 2 implies that $q = q'$. $\square$

Corollary. *Let P be a finite biprefix. If the syntactic ts $X = (Q, S)$ of P^+ is a transformation monoid, then X is injective.*

Proof. By the above $X = X_1$ is injective. $\square$

We now show that any injective ts $X = (Q, S)$ divides the syntactic ts of a finite biprefix. We first recall some results from [7].

Let $\mathcal{A} = (Q, \Sigma)$ be a Σ-automaton, with $Q = \{1, \ldots, n\}$. Let $A = \{a\} \cup \Sigma$ with $a \notin \Sigma$. The prefix $P(\mathcal{A}) = \{a^{2^i}\sigma a^{2^n - 2^{i\sigma}} \mid 1 \leq i \leq n,\ \sigma \in \Sigma,\ i\sigma \neq \emptyset\}$ is called the *Pin Code* of $\mathcal{A}$.

The following appears in [7].

Theorem 4. *Let X be the ts of $\mathcal{A}$, and let Y be the syntactic ts of $P(\mathcal{A})^+$. Then X divides Y.*

Lemma 5. *The ts X of $\mathcal{A}$ is an injective ts if and only if $P(\mathcal{A})$ is a biprefix.*

Proof. First note that X is injective if and only if each $\sigma \in \Sigma$ induces an injective function on Q. Furthermore $a^{2^i}\sigma a^{2^n-2^{i\sigma}}$ is a suffix of $a^{2^j}\tau a^{2^n-2^{j\tau}}$ if and only if $\sigma = \tau$, $i \leqslant j$ and $i\sigma = j\sigma$. Therefore, X is injective if and only if $P(\mathscr{A})$ is a biprefix. $\square$

Lemma 5 was also observed by Pin (private communication).

Theorem 6. *If $\mathscr{A}$ is an injective automaton, then so is the minimal automaton of $P(\mathscr{A})^+$.*

Proof. Let $Y = (P, A)$ be the minimal automaton of $P(\mathscr{A})^+$. In [7] it is shown that

$$P = \{q_j \mid -m \leqslant j \leqslant 2^n\} \quad \text{where } m = \max_{\substack{\sigma \in \Sigma \\ i\sigma \neq \emptyset}} (2^n - 2^{i\sigma})$$

and

$$q_j = (a^j)^{-1} P(\mathscr{A})^+, \quad 0 \leqslant j \leqslant 2^n$$

and

$$q_{-j} = a^j P(\mathscr{A})^*, \quad 1 \leqslant j \leqslant m.$$

Furthermore

$$q_j a = \begin{cases} q_{j+1} & \text{if } j+1 \leqslant 2^n, \\ \text{undefined} & \text{otherwise} \end{cases}$$

and if $\sigma \in \Sigma$,

$$q_j\sigma = \begin{cases} q - 2^n + 2^{i\sigma} & \text{if } i\sigma \neq \emptyset \text{ and } j = 2^i, \\ \text{undefined} & \text{otherwise.} \end{cases}$$

It follows easily from these results, that if each $\sigma \in \Sigma$ induces an injective function on Q, then each letter of A induces an injective function on P. $\square$

Corollary 1. *Every injective ts divides an injective ts which is the syntactic ts of a finite biprefix.*

Recall that a variety of finite semigroups is a collection of finite semigroups closed under division and direct product. A variety of rational languages is a collection of rational languages closed under union, complementation, quotients and inverse morphism. Eilenberg's Theorem sets up a one to one correspondence between varieties of finite semigroups and varieties of rational languages. See [1] and [3] for details.

Following Pin [7] we say that a variety $\mathscr{V}$ of rational languages is described by a class $\mathscr{C}$ of prefixes if $\mathscr{V}$ is the smallest variety containing P^+ for all $P \in \mathscr{C}$. Let $\mathscr{I}n$ be the variety of rational languages corresponding to the variety $\underline{In}$ of semigroups generated by inverse semigroups.

Corollary 2. *$\mathscr{I}n$ is described by its finite biprefixes.*

Proof. Let $S \in \underline{In}$. Then S divides an inverse semigroup T. As is well known, T has a faithful representation by injective functions on T The results now follow from Corollary 1 and Eilenberg's Theorem. $\square$

If X is a finite subset of A^+, define the complexity Xc of X to be the complexity of the semigroup $S(X^+)$. See [12] for an exposition of complexity theory.

The following is proved in [4].

Theorem 7. *The complexity of X is less than or equal to* card(X).

If X is a biprefix we have

Theorem 8. *Let X be a finite biprefix. Then $Xc \leqslant 1$.*

Proof. Let $Y = (P, T)$ be the syntactic ts of X^+. By Theorem 3, Y is locally injective. In particular, the transformation monoid $\bar{2}^{\cdot}$ does not divide Y. Recall that $\bar{2}^{\cdot}$ has two states and the identity map and the two constant maps as transformations. It follows from the results of [1, Chapter 4], that $Y\mathscr{C} \leqslant 1$. Since $Yc = Tc$, the theorem is proved. $\square$

3. Some open problems

(1) Find necessary and sufficient conditions for a finite prefix P to be such that the syntactic ts of P^+ is locally injective.

Any finite biprefix and any finite very pure prefix is locally injective. J.E. Pin (private communication) has given the following construction of locally injective finite prefixes. Let A and B be alphabets and let $f: A^+ \to B^+$ be a non-trivial morphism such that Af is a complete biprefix.

Let P be a finite very pure prefix. Then Pf is a locally injective prefix which is neither a biprefix nor very pure. Are all finite locally injective prefixes which are not very pure nor biprefix obtained this way?

(2) Let LIn be the variety of semigroups S such that eSe divides an inverse semigroup for all idempotents $e \in S$. Is LIn described by its finite prefixes?

The author has constructed an automaton $\mathscr{A} = (Q, \Sigma)$ such that the ts of $\mathscr{A}$ is locally injective, but the syntactic ts of $P(\mathscr{A})^-$ is not locally injective. A positive solution to this problem would be useful in applying the theory of prefixes to the complexity theory of ts where locally injective ts's play an important role.

References

[1] S. Eilenberg, *Automata, Languages and Machines, Vol. B* (Academic Press, New York, 1976).

[2] M. Keenan and G. Lallement, On certain codes admitting inverse semigroups as syntactic monoids, *Semigroup Forum* **8** (1974) 312–331.
[3] G. Lallement, *Semigroups and Combinatorial Applications* (Wiley, New York, 1979).
[4] E. Le Rest and S.W. Margolis, Complexity of finitely generated submonoids of a free monoid, to appear.
[5] S.W. Margolis and J.E. Pin, On varieties of rational languages and variable length codes, II, to appear.
[6] J.E. Pin, Sur le monoide syntactique de L^* lorsque L est un langage fini, Theoret. Comput. Sci., to appear.
[7] J.E. Pin, On varieties of rational languages and variable length codes, I, *J. Pure Appl. Alg.* **23** (1982) 169–196.
[8] C. Reutenauer, Une topologie du monoide libre, *Semigroup Forum* **18** (1979) 33–49.
[9] C. Reutenauer, Sur mon article "Une topologie du monoide libre", *Semigroup Forum* **22** (1981) 93–95.
[10] M.P. Schutzenberger, Une théorie algèbrique du codage, *C.R. Acad. Sci. Paris* **242** (1956) 862–864.
[11] M.P. Schutzenberger, Sur le produit de concaténation non ambigu, *Semigroup Forum* **13** (1976) 47–75.
[12] B. Tilson, Complexity of semigroups and morphisms, in: S. Eilenberg, *Automata, Languages and Machines, Vol. B* (Academic Press, New York, 1976).