FREE SUBSEMIGROUPS OF A FREE SEMIGROUP

E. K. Blum

1. INTRODUCTION

Let Σ be an alphabet, and let $F(\Sigma)$ be the free semigroup of words over Σ. (See [2] for definitions.) We denote the empty word by e. If $G \subset F(\Sigma)$ and $H \subset F(\Sigma)$, then $GH = \{uv \mid u \in G,\ v \in H\}$. Further, $G^1 = G$, $G^n = GG^{n-1}$, and $G^0 = \{e\}$. We define

$$G^* = \bigcup_{i=0}^{\infty} G^i \quad \text{and} \quad G^{\infty} = \bigcup_{i=1}^{\infty} G^i.$$

Clearly, G^∞ is precisely the subsemigroup of $F(\Sigma)$ generated by G. In this paper, we give a necessary and sufficient condition on G in order that G^∞ be free. Our condition differs from those given in [1] and [3] in that it can be used to obtain a constructive procedure for determining whether G^∞ is free, in the case where G is finite.

2. SUBSEMIGROUPS OF $\overline{F}(\Sigma)$

Let $\overline{F}(\Sigma) = F(\Sigma) - \{e\}$; that is, let $\overline{F}(\Sigma)$ be the subsemigroup of all nonempty words. If $E \subset \overline{F}(\Sigma)$ and $u \in \overline{F}(\Sigma)$, then by an *E-factorization* of u we shall mean a representation of u as a product of factors each of which is in E; that is,

$$u = v_1 v_2 \cdots v_n \qquad (v_i \in E,\ 1 \leq i \leq n).$$

If every $u \in E^\infty$ has a unique E-factorization, then E^∞ is a free subsemigroup of $\overline{F}(\Sigma)$ with E as its unique irreducible generating set, and conversely [2].

We denote the length of any $u \in F(\Sigma)$ by $|u|$.

Definition 1. Let $E \subset \overline{F}(\Sigma)$. A sequence $\{t_0, t_1, \cdots, t_n, e\}$ $(n \geq 0,\ t_i \in F(\Sigma))$ is an *E-chain* if and only if for $1 \leq i \leq n$ there exist $w_i \in E$ and $v_i \in E^*$ such that $w_i = t_i v_i t_{i+1}$, where $t_{n+1} = e$, and for $i = 0$ either there exists $v_0 \in E^*$ such that $t_0 = v_0 t_1$ or there exist $w_0 \in E$ and $v_0 \in E^*$ such that $w_0 = t_0 v_0 t_1$. We call w_i the *ith link* of the chain. If $t_0 = v_0 t_1$, then $v_0 t_1$ is the *zeroth link* of the chain. The word t_{i+1} is called a *successor* of t_i.

Remark. Obviously, if $\{t_1, t_2, \cdots, t_n, e\}$ $(n \geq 1)$ is an E-chain and $t_0 \in F(\Sigma)$ is such that there exist $w_0 \in E$ and $v \in E^*$ such that either $t_0 = vt_1$ or $w_0 = t_0 v t_1$, then either $\{t_0, t_1, t_2, \cdots, t_n, e\}$ or $\{t_0, t_2, \cdots, t_n, e\}$ is also an E-chain, depending on whether $w_1 = t_1 v_1 t_2$ or $t_1 = v_1 t_2$. In the latter case, the zeroth link from t_0 to t_2 is $w_0 = t_0 v' t_2$ or $t_0 = v' t_2$, where $v' = vv_1$. If n = 1, the chain is simply $\{t_0, e\}$. Note also that if $t_0 = e$, then $\{e, e\}$ is a (trivial) E-chain. Likewise, if $t_0 = v \in E^*$, then $\{t_0, e\}$ is a (trivial) E-chain.

Received September 6, 1964.
This research was partially supported by N.S.F. Grant GP 2993.

LEMMA 1. *If* $v = t_0 v'$ *and* $v \in E^*$, $v' \in E^*$, $t_0 \in F(\Sigma)$, *then there exists an* E-*chain starting with* t_0 *as the first word in the chain.* ($E \subset \overline{F}(\Sigma)$.)

Proof. By induction on $|v|$, $v \in E^*$.

Remark. If $t_0 = e$, then the trivial E-chain $\{e, e\}$ serves to prove the lemma.

Basis. $|v| = 0$. Since $t_0 = e$, the trivial E-chain $\{e, e\}$ establishes the basis.

Induction step. Assume the lemma holds for all $v \in E^*$ of length less than k, and suppose (1) $v = t_0 v'$ with $|v| = k > 0$. In view of the above remark, we may take $t_0 \neq e$, which means $|v'| < k$. Since $v \neq e$, there is a word $w_0 \in E$ such that $v = w_0 v_1$, with $v_1 \in E^*$. Substituting for v in (1), we obtain (2) $w_0 v_1 = t_0 v'$. From (2) we see that either

$$(a)\ w_0 = t_0 t_1 \quad \text{or} \quad (b)\ t_0 = w_0 t_1, \quad t_1 \in F(\Sigma).$$

In case (a), substitution in (2) and left-cancellation of t_0 gives (3) $t_1 v_1 = v'$. In case (b), we obtain (4) $v_1 = t_1 v'$. Both of the equations (3) and (4) are of the form considered in the lemma, and both $|v'| < k$ and $|v_1| < k$. Hence, in either case, the induction hypothesis applies and there is an E-chain starting with t_1, say $\{t_1, t_2, \cdots, t_n, e\}$ $(n \geq 1)$. We now obtain the desired E-chain as either $\{t_0, t_1, t_2, \cdots, t_n, e\}$ $(n \geq 1)$, in case (a) by using w_0 as the zeroth link, and in case (b) by using $w_0 t_1$ as the zeroth link or as $\{t_0, t_2, \cdots, t_n, e\}$ according to the remark following Definition 1. This completes the proof.

Definition 2. Let E be a subset of $\overline{F}(\Sigma)$. A distinct pair $\{w, w'\} \subset E$ is an E-*couple* if $w = w' t_0$ $(t_0 \in F(\Sigma))$ and there exists an E-chain $\{t_0, t_1, \cdots, t_n, e\}$ $(n \geq 0)$ starting with t_0.

Remark. If $w = w' v$ and $v \in E^*$, then $\{v, e\}$ is an E-chain (with $t_0 = v$) that serves to make $\{w, w'\}$ an E-couple.

THEOREM 1. *Let* $E \subset \overline{F}(\Sigma)$. *A necessary and sufficient condition that* E^∞ *be a free subsemigroup of* $\overline{F}(\Sigma)$ *with* E *as its unique irreducible generating set is that* E *contains no* E-*couple.*

Proof. Sufficiency. Suppose E^∞ is not free. Then for some $u \in E^\infty$ there must exist two distinct E-factorizations such that

$$u = wv = w'v', \quad \text{where } w \neq w', \quad w, w' \in E, \quad v, v' \in E^\infty.$$

Either $w = w't_0$ or $w' = wt_0$ $(t_0 \in F(\Sigma))$. In the first case, $t_0 v = v'$, and in the second, $v = t_0 v'$. By Lemma 1, there exists an E-chain starting with t_0. Hence, $\{w, w'\}$ is an E-couple.

Necessity. Suppose E contains an E-couple $\{w, w'\}$. In the notation of Definitions 1 and 2, we have (1) $w = w' t_0$, either (2.1) $t_0 = v_0 t_1$ or (2.2) $w_0 = t_0 v_0 t_1$, and (3) $w_i = t_i v_i t_{i+1}$ for $1 \leq i \leq n$, with $t_{n+1} = e$. From (3) it follows that

$$w_1 v_2 w_3 v_4 \cdots = t_1 v_1 w_2 v_3 w_4 \cdots,$$

since both sides are equal to $(t_1 v_1)(t_2 v_2)(t_3 v_3)(t_4 v_4) \cdots$. From (1) and (2.1), multiplying both sides by $w' v_0$, we obtain

$$w' v_0 w_1 v_2 w_3 v_4 \cdots = w v_1 w_2 v_3 w_4 \cdots .$$

Alternately, from (1) and (2.2), multiplying both sides by $w v_0$, we obtain

$$w v_0 w_1 v_2 w_3 v_4 \cdots = w' w_0 v_1 w_2 v_3 w_4 \cdots .$$

In either case, since $w \neq w'$, we obtain a word in E^∞ that has two distinct E-factorizations. (If $n = 0$, then either $w'v_0 = w$ or $w v_0 = w' w_0$.) Hence, E^∞ is not a free subsemigroup with E as its unique irreducible generating set. This completes the proof.

It is known that every finitely generated semigroup has an irreducible generating set and that this holds also for any subsemigroup of $\overline{F}(\Sigma)$. (Indeed, if G is a cancellation semigroup without identity in which every element has only a finite number of left-factors, then G has a unique irreducible generating set.) As a consequence, the following corollary of Theorem 1 follows immediately.

COROLLARY. *Let* G *be a subsemigroup of* $\overline{F}(\Sigma)$ *with* E *as its unique irreducible generating set.* G *is a free semigroup if and only if* E *has no* E*-couples.*

3. SUBSEMIGROUPS OF $\overline{F}(\Sigma)$ WITH FINITE GENERATING SETS

The condition on E-couples has the consequence that for finite E there exists an effective (finite) procedure for determining whether E^∞ is free.

THEOREM 2. *Let* $E = \{w_1, \cdots, w_m\}$ $(m \geq 1)$ *be a subset of* $\overline{F}(\Sigma)$. *There exists an effective decision procedure for determining whether or not an arbitrary pair* $\{w_i, w_j\} \subset E$ *is an* E*-couple.*

Proof. We demonstrate the procedure for $\{w_1, w_2\}$, the same steps being applicable to any pair $\{w_i, w_j\}$ in E.

First, determine whether either word is a left-factor of the other. If neither is the case, then $\{w_1, w_2\}$ is not an E-couple. If $w_1 = w_2 t_0$, say, we must determine whether t_0 is the beginning of an E-chain. This can be done in a finite number of steps as follows.

Consider the set of successors of t_0 (Definition 1). This set is finite or empty, since any successor t_1 must be such that either (1) $w_j = t_0 v_0 t_1$ or (2) $t_0 = v_0 t_1$, with $w_j \in E$ and $v_0 \in E^*$. We find all successors of t_0 by testing for all possible instances of cases (1) and (2). To determine whether (1) obtains, we consider each $w_j \in E$ $(1 \leq j \leq m)$ and test whether t_0 is a left-factor of w_j. If $w_j = t_0 u$, we determine all factorizations of u of the form vt with $v \in E^*$. This can be done in a finite number of steps. If $v_0 t_1$ is any such factorization, then $w_j = t_0 v_0 t_1$, and t_1 is a successor of t_0. To determine whether (2) obtains, we consider each of the right-factors of t_0. If the corresponding left-factor is in E^*, then we have an instance of case 2.

To each successor t_1 of t_0 found in this way we apply the same procedure and thereby determine all successors of t_1. Continuing in this manner, we generate a rooted tree, starting with t_0, and branching at each node t_i to all possible successors of t_i, thereby generating new nodes one level higher. Since each node t_j must be a right-factor of some word in E or a right-factor of t_0, there are at most a finite number, say N, of distinct words t_j. We assert that it suffices to generate at most the Nth level of the tree. For suppose there exists an E-chain starting with t_0, say $\{t_0, t_1, \cdots, t_n, e\}$. If some t_i is t_{i+r-1} $(r > 1)$, then

$$\{t_0, t_1, \cdots, t_i, t_{i+r}, \cdots, t_n, e\}$$

is also an E-chain. Thus, if there exists an E-chain starting with t_0, then there exists one, $\{t_0, t_1, \cdots, t_n, e\}$, with $n < N$. Any such E-chain corresponds to some path in the tree generated by the above procedure at the Nth level or earlier.

REFERENCES

1. P. M. Cohn, *On subsemigroups of free semigroups*, Proc. Amer. Math. Soc. 63 (1962), 347-351.

2. E. S. Lyapin, *Semigroups*, Translations of Mathematical Monographs, vol. 3, Amer. Math. Soc., Providence, R.I., 1963.

3. R. C. Lyndon and M. P. Schützenberger, *The equation* $a^M = b^N c^P$ *in a free group*, Michigan Math. J. 9 (1962), 289-298.

Wesleyan University
Middletown, Connecticut