ON FREE MONOIDS PARTIALLY ORDERED BY EMBEDDING

by

Leonard H. Haines

Memorandum No. ERL-M219

29 August 1967

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

On Free Monoids Partially Ordered by Embedding*

Leonard H. Haines

Department of Electrical Engineering and
Computer Sciences and Electronics Research Laboratory
University of California, Berkeley, California

ABSTRACT

A combinatorial theorem about finitely generated free monoids is proved and used to show that the set of all subsequences (or super-sequences) of any set of words in a finite alphabet is a regular event.

*The research reported herein was supported wholly by the National Science Foundation under Grant GP-6945.

INTRODUCTION

Let Σ^* be the free monoid with null word ϵ generated by a finite alphabet Σ. Let $\leq$ partially order Σ^* by embedding (i.e., $x \leq y$ iff $x = x_1 x_2 \ldots x_n$ and $y = y_1 x_1 y_2 x_2 \ldots y_n x_n y_{n+1}$ for some integer n where x_i and y_j are in Σ^* for $1 \leq i < j \leq n+1$).

THEOREM 1. Each set of pairwise incomparable elements of Σ^* is finite.[1]

For any $A \subset \Sigma^*$ define

$$\tilde{A} = \{x \text{ in } \Sigma^* : y \leq x \text{ for some } y \text{ in } A\}$$

and

$$\underset{\sim}{A} = \{x \text{ in } \Sigma^* : x \leq y \text{ for some } y \text{ in } A\}$$

THEOREM 2. Let $A \subset \Sigma^*$. Then there exist finite subsets F and G of Σ^* such that $\tilde{A} = \tilde{F}$ and $\underset{\sim}{A} = \Sigma^* - \tilde{G}$.

THEOREM 3. $\tilde{A}$ and $\underset{\sim}{A}$ are regular sets for any $A \subset \Sigma^*$.

In Section 2 we will show that Theorem 1 $\Rightarrow$ Theorem 2 $\Rightarrow$ Theorem 3. For ease of reading the proof of Theorem 1 is deferred until Section 3.

An easy corollary of Theorem 1 is a well known result of König[2].

COROLLARY (König). Each set of pairwise incomparable elements of $(N^k, \leq)$ is finite (where N^k, the set of k-tuples over the non-negative integers N, is partially ordered so that $(u_1, u_2, \ldots, u_k) \leq (v_1, v_2, \ldots, v_k)$ iff $u_i \leq v_i$ for $1 \leq i \leq k$).

Note that Theorem 1 fails if Σ^* is partially ordered by sub-words, i.e., if $\leq_1$ is defined so that $x \leq_1 y$ iff $y = y_1 x y_2$ for some y_1 and y_2 in Σ^* then, for a and b in Σ, $\{a b^n a : n \geq 1\}$ is an infinite set of pairwise incomparable elements of $(\Sigma^*, \leq_1)$. Similar counter examples exist for $(\Sigma^*, \leq_k)$, where $x \leq_k y$ iff $x = x_1 x_2 \ldots x_k$ and $y = y_1 x_1 y_2 x_2 \ldots y_k x_k y_{k+1}$ for some x_i and y_j in Σ^* $(1 \leq i < j \leq k+1)$. Any necessary and sufficient conditions on partial orderings which insure Theorem 1 must exclude $(\Sigma^*, \leq_k)$ which shares many formal properties with $(\Sigma^*, \leq)$.

Theorem 3 is unexpected. One might suppose that $\underset{\sim}{A}$ can be non-recursive for suitably chosen A (e.g. A the domain of a partial recursive function defined by a Turing Machine which accepts an input word w iff every subsequence of w satisfies an appropriate predicate. Evidently no such predicate exists).

The proof of Theorem 3 (and therefore Theorem 2) is necessarily non-constructive for recursively enumerable A. This is clear since A is empty iff $\tilde{A}$ is empty iff $\underset{\sim}{A}$ is empty but the question of whether a set is empty is undecidable for arbitrary recursively enumerable sets and decidable for arbitrary regular sets.[2] Indeed, for the very same reason, given a context-sensitive grammar G one cannot effectively construct the regular events which represent $\widetilde{L(G)}$ and $\underset{\sim}{L(G)}$. Given a context-free grammar G it is simple exercise to construct context-free grammars G_1 and G_2 such that $L(G_1) = \widetilde{L(G)}$ and $L(G_2) = \underset{\sim}{L(G)}$. Whether G_1 and G_2 can be effectively transformed into the regular events (or finite automata or right linear grammars) which specify $\widetilde{L(G)}$ and $\underset{\sim}{L(G)}$ is an interesting open problem. Ullian[3] has shown that one cannot effectively transform a context-free grammar G which generates a regular language into a regular event which represents $L(G)$. In fact, one cannot effectively determine whether $L(G)$ is Σ^* or $\Sigma^* - \{w\}$ for some non- ϵ word w even when these are known to be the only possibilities.

PROOF OF THEOREMS 2 AND 3

THEOREM 2a. Let $A \subset \Sigma^*$. Then there exists a finite subset F of Σ^* such that $\tilde{A} = \tilde{F}$.

<u>Proof.</u> Let F be the set of all minimal elements of A. Clearly $\tilde{A} = \tilde{F}$. By Theorem 1 F must be finite.

THEOREM 2b. Let $A \subset \Sigma^*$. Then there exists a finite subset G of Σ^* such that $\underset{\sim}{A} = \Sigma^* - G$.

Proof. Let $B = \Sigma^* - \underset{\sim}{A}$. By definition $B \subset \tilde{B}$. Now suppose that $\tilde{B} \not\subset B$, i.e., suppose that there is a word x in $\tilde{B} \cap \underset{\sim}{A}$. Then since x is in $\tilde{B}$, $x \geq y$ for some y in B. On the other hand, since x is also in $\underset{\sim}{A}$, y is also in $\underset{\approx}{A} = \underset{\sim}{A} = \Sigma^* - B$ which is absurd. Hence $B = \tilde{B}$ and therefore by Theorem 2a $B = \tilde{G}$ for some finite set G so that $\underset{\sim}{A} = \Sigma^* - G$.

Proof of Theorem 3. For any word w in Σ^* $\tilde{w}$ is obviously regular since

$$\tilde{w} = \Sigma^* w_1 \Sigma^* w_2 \ldots \Sigma^* w_n \Sigma^*$$

where $w = w_1 w_2 \ldots w_n$ for w_i in $\Sigma \cup \{\epsilon\}$, $1 \leq i \leq n$. Since a finite union of regular sets is regular, $\tilde{W} = \cup \{\tilde{w} : w \text{ in } W\}$ is regular for any finite subset W of Σ^*. Now if F and G are as in Theorem 2 then $\tilde{A} = \tilde{F}$ and $\tilde{G}$ are regular as is $\underset{\sim}{A} = \Sigma^* - \tilde{G}$ since the complement of a regular set is regular.

PROOF OF THEOREM 1 [3]

Lemma. If Theorem 1 holds for an alphabet Σ then every infinite subset of Σ^* possesses an infinite chain.

Proof. Let A be an infinite subset of Σ^* and suppose that every chain in A is finite. The totality of maximum elements of maximal chains in A is identical with the maximum elements of A and is therefore, by hypothesis, finite. Since A is infinite, infinitely many distinct chains have the same maximum element u. But then infinitely many and therefore arbitrarily long elements of Σ^* precede u, contradicting the definition of $\leq$.

The proof of Theorem 1 is by induction on the size of Σ. For 1 - letter alphabets the theorem is trivial. Suppose that Theorem 1 holds for all n-letter alphabets and fails for an $n+1$ letter alphabet Σ.

For each infinite set of pairwise incomparable elements $Y = \{y_1, y_2, \dots\}$ of Σ^* there is shortest x in Σ^* such that $x \not\leq y_i$ holds for all i. Without loss of generality we may suppose that Y is chosen so that x is of minimal length. Clearly $x \neq \epsilon$.

Let

$$x = x_1 x_2 \dots x_k, \quad x_j \text{ in } \Sigma, \quad 1 \leq j \leq k.$$

If $k = 1$ then y_i is in $(\Sigma - x_1)^*$ for all $i \geq 1$ which contradicts the induction hypothesis. Because of the choice of x,

$$x_1 x_2 \dots x_{k-1} \leq y_i$$

holds for all but finitely many i and therefore by relabeling subscripts we may assume it holds for all $i \geq 1$. Hence for each $i \geq 1$ there exist unique words $y_{i1}, y_{i2}, \dots y_{ik}$ such that

$$y_i = y_{i1} x_1 y_{i2} x_2 \dots y_{i\,k-1} x_{k-1} y_{ik}$$

and $x_j \not\leq y_{ij}$ holds for $1 \leq j < k$. Furthermore the choice of x guarantees that $x_k \not\leq y_{ik}$ holds for all $i \geq 1$.

We now assert that there are infinite index sets $N_1, N_2, \dots, N_k$ such that $N_j \supset N_{j+1}$ $(1 \leq j < k)$ and $y_{pj} \leq y_{qj}$ whenever p and q are in N_j $(1 \leq j \leq k)$ and $p < q$. Let $N_o = \{i : i \geq 1\}$. We will establish the existence of N_j from the existence of N_{j-1}, $1 \leq j \leq k$.

Let

$$Y_j = \{y_{ij} : i \text{ in } N_{j-1}\}.$$

If Y_j is finite then at least one of the sets $\{i \text{ in } N_{j-1} : y_{ij} = w\}$ is infinite for some fixed word w and we may choose N_i to be any such infinite set. Alternatively, if Y_j is infinite, the induction hypothesis (applicable since $Y_j \subset (\Sigma - x_j)^*$) and the lemma imply that Y_j possesses an infinite chain $y_{s_1 j} < y_{s_2 j} < \dots$. Now if $t_1, t_2, \dots$ is any infinite strictly increasing subsequence of $s_1, s_2, \dots$ then we may choose $N_j = \{t_i : i \geq 1\}$. Hence the assertion is valid.

But then if $p < q$ are in N_k then p and q are also in N_j $(1 \leq j < k)$ so that $y_{pj} \leq y_{qj}$ $(1 \leq j \leq k)$ and therefore

$$y_p = y_{p1} x_1 y_{p2} x_2 \dots y_{p\,k-1} x_{k-1} y_{pk}$$
$$\leq y_{q1} x_1 y_{q2} x_2 \dots y_{q\,k-1} x_{k-1} y_{qk} = y_q,$$

a contradiction which establishes the theorem.

FOOTNOTES

1. Theorem 1 can be reformulated as an amusing combinatorial property of real numbers: no matter how one partitions an infinite n-ary expansion of any real number into blocks of finite length one block is necessarily a subsequence of another.

2. See Ginsberg[1] for the definition and properties of regular sets, regular events, context-free and context-sensitive grammars.

3. I am indebted to Robert Solovay for his help in extending a previous proof of Theorem 1 beyond the special case of 3-letter alphabets.

REFERENCES

1. S. Ginsburg, Mathematical Theory of Context-free Languages, McGraw Hill, 1966.

2. D. König, Theorie der Endlichen und Unendlichen Graphen, Chelsea Publishing Company, New York, 1950.

3. J. Ullian, "Partial Algorithm Problems for Context Free Languages," System Development Corporation Report TM-738/027/00, October 10, 1966.