Frontiers of one-letter languages

Stephen L. Bloom

Introduction

In a talk titled "Infinite words" given in the spring of '83, Dana Scott introduced the notion of a convergent sequence of "word" (i.e. elements of a finitely generated free monoid). A sequence is convergent if, for every regular set R of words, all but finitely many of the terms of the sequence belong to R or all but finitely many belong to the complement of R. Scott stated a number of properties of the collection of (equivalence classes of) convergent sequences, some of which showed that this structure had been known earlier in the guise of a free "profinite" monoid. This monoid has a natural topology and contains a copy of the original free monoid (as a certain discrete subspace). In the case that the words are elements of the one-generated free monoid (i.e. the additive monoid N of the nonnegative integers) an explicit description of all convergent sequences was stated (and is derived here). Further, he posed several problems connected with this structure and the current paper addresses one of these.

Scott asked whether an investigation of the "frontiers" of languages would lead to a useful classification of languages. In an effort to answer this question, we looked at the simplest case: frontiers of subsets of N. An explicit description of these frontiers has been obtained. It is seen that the cardinality of the frontier of a subset of N cannot distinguish regular from nonregular sets, and a necessary and sufficient condition is given that a subset of N have an uncountable frontier.

In order to obtain these facts we derived a number of the facts mentioned by Scott in his talk. We have tried to give very elementary proofs and keep the paper self-contained.

Preliminaries

N will denote both the set of nonnegative (or "natural") numbers as well as the additive monoid $(N, +, 0)$. An infinite sequence will be denoted by diagonal brackets:

$$\langle x\rangle = \langle x_1, x_2, \ldots\rangle;$$

the *terms* of $\langle x\rangle$ are the elements x_n, $n>0$.

The value of a function f on an argument x will be written as xf or $f(x)$.

For positive integers k and l, the 1-generated finite monoid which is the quotient of N by the least monoid congruence identifying $k+l$ and k is denoted $N_{k,l}$. When $k=l=n!$, this monoid will be denoted M_n. One may assume that the elements of $N_{k,l}$ are the integers

$$0, 1, \ldots, k, k+1, \ldots, k+l-1.$$

The closure of a subset A of a topological space is denoted A^-, and the *frontier* of A is $A^- - A$.

In this paper, we will use the word "uncountable" to mean the cardinality of the powerset of the natural numbers.

1. Convergent sequences

Let $\langle x \rangle$ and $\langle y \rangle$ be infinite sequences of nonnegative integers. For a subset S of N, the nonnegative integers, define $\langle x \rangle$ to be "S-equivalent" to $\langle y \rangle$ if, all but finitely many of the terms of $\langle x \rangle$ belong to S iff all but finitely many of the terms of $\langle y \rangle$ do.

1.1. Definition. $\langle x \rangle$ is equivalent to $\langle y \rangle$, written $\langle x \rangle \sim \langle y \rangle$, if for every *regular* set R, $\langle x \rangle$ is R-equivalent to $\langle y \rangle$.

It is easy to verify that $\sim$ is indeed an equivalence relation on the infinite sequences of natural numbes. The $\sim$ – equivalence class of $\langle x \rangle$ will be written $\langle x \rangle^\sim$.

1.2. Definition. An infinite sequence is *convergent* if, for each regular set R, either all but finitely many terms of $\langle x \rangle$ belong to R (we will say $\langle x \rangle$ is "eventually in R") or $\langle x \rangle$ is eventually in $N \setminus R$, the complement of R.

We let *Conv* denote the set of all convergent sequences. Note that if $\langle x \rangle$ is convergent and $\langle y \rangle$ is equivalent to $\langle x \rangle$, then $\langle y \rangle$ is also convergent.

The set of equivalence classes of convergent sequences is denoted N^-. This set has both a monoid structure and a topological structure, as will be shown below, and is one of the more fascinating objects in the mathematical universe. Our goal in this first section is to give an explicit representation of the members of N^-.

We will say that the sequence $\langle y \rangle$ is a subsequence of the sequence $\langle x \rangle$ if there is a strictly increasing function f: $N \to N$ such that for each $n \in N$,

$$y_n = x_{nf}.$$

1.3. Proposition. If $\langle y \rangle$ is a subsequence of the convergent sequence $\langle x \rangle$, then $\langle y \rangle$ is equivalent to $\langle x \rangle$ and is therefore convergent.

A sequence of integers is *eventually increasing* if for each number b, $x_n > b$ for all but finitely many n; similarly a sequence $\langle x \rangle$ is *eventually constant* with value b if $x_n = b$ for all but finitely many n.

1.4. Proposition. If $\langle x \rangle$ is convergent, either $\langle x \rangle$ is eventually constant or eventually increasing.

Proof. Let R_b denote the regular set of integers greater than b. Then either $\langle x \rangle$ is eventually in R_b for every b, in which case $\langle x \rangle$ is eventually increasing, or not. If not, since $\langle x\rangle$ is convergent, $\langle x\rangle$ is eventually in the finite union of the singleton sets $\{0\}\cup\{1\}\cup\ldots\cup\{b\}$, for some b. We now apply the following lemma.

1.5. Lemma. If $\langle x\rangle$ is convergent and $\langle x\rangle$ is eventually in a finite union of regular sets $A1\cup\ldots\cup An$, then for some i, $\langle x\rangle$ is eventually in Ai.

By the Lemma 1.5, $\langle x\rangle$ is eventually constant, proving 1.4. The easy proof of 1.5 is omitted.

1.6. Proposition. Let $\langle x\rangle$ be a convergent sequence. Then either there is a constant sequence $\langle y\rangle$ or a strictly increasing sequence $\langle y\rangle$ with $\langle x\rangle$ equivalent to $\langle y\rangle$.

Proof. By 1.4, $\langle x\rangle$ is either eventually constant, in which case $\langle x\rangle$ has a constant subsequence, or $\langle x\rangle$ is eventually increasing, in which case $\langle x\rangle$ has a strictly increasing subsequence. By 1.3, the proof is complete.

We will characterize those strictly increasing sequences which are convergent. But first we give an alternate formulation of the notion of convergent sequence.

1.7. Proposition. The sequence of integers $\langle x\rangle$ is convergent iff

(1.7a) for each homomorphism h from the additive monoid $(N, +, 0)$ to a finite monoid M, the sequence

$$\langle xh\rangle = \langle x_1 h, x_2 h, \ldots\rangle$$

is eventually constant.

Proof. Suppose first that $\langle x\rangle$ is convergent and that $h\colon N\to M$ is a monoid homorphism. If M consists of the elements $m1, \ldots, mk$, then the sets

$$Ri = h^{-1}(mi)$$

are disjoint regular sets and $\langle x\rangle$ is eventually in the union $R1\cup\ldots\cup Rk=N$. By lemma 1.5, $\langle x\rangle$ is eventually in one Ri; i.e. $\langle xh\rangle$ is eventually constant.

Now suppose that (1.7a) holds. Let R be any regular set and recall the following fundamental fact:

any regular subset R of N may be written as

$$R = h^{-1}(X)$$

for some monoid homomorphism $h\colon N\to M$, M finite, and some $X\subset M$.

Thus, $\langle x\rangle$ is eventually in R if $\langle xh\rangle$ is eventually in X; otherwise, $\langle x\rangle$ is eventually in the complement of R.

The preceding proposition may be rephrased: a convergent sequence is one whose terms eventually cannot be distinguished by a finite automaton.

The next fact is proved in exactly the same way.

1.8. Proposition. If $\langle x\rangle$ and $\langle y\rangle$ are convergent sequences, then $\langle x\rangle$ is equivalent to $\langle y\rangle$ iff for every monoid homomorphism $h\colon N\to M$, M finite, $\langle xh\rangle$ and $\langle yh\rangle$ are eventually equal.

1.9. Corollary. If $\langle x\rangle$ and $\langle y\rangle$ are convergent, so is "$\langle x\rangle+\langle y\rangle$", where the "sum" is obtained by pointwise addition; further, if $\langle x\rangle\sim\langle x'\rangle$ and $\langle y\rangle\sim\langle y'\rangle$, then $\langle x\rangle+\langle y\rangle\sim\langle x'\rangle+\langle y'\rangle$.

Both facts follow from 1.7 and 1.8.

1.10. Corollary. N^- is a monoid, where the operation is that in 1.9 and the identity element is the equivalence class of the constant zero sequence.

The next proposition is quite useful.

1.11. Proposition. Suppose that $\langle x \rangle$ is a strictly increasing sequence. Then $\langle x \rangle$ is convergent iff, for each $n>0$,

$$x_m \equiv x_p \pmod{n!} \tag{1.11a}$$

for all but finitely many m and p. If $\langle x \rangle$ and $\langle y \rangle$ are strictly increasing convergent sequences, $\langle x \rangle \sim \langle y \rangle$ iff, for each $n>0$,

$$x_m \equiv y_m \pmod{n!} \tag{1.11b}$$

for all but finitely many m.

Proof. If $\langle x \rangle$ is convergent, then (1.11a) holds, by 1.7 applied to the canonical homomorphism $N \to Z/n!$, where $Z/n!$ denotes the ring of integers modulo $n!$. Conversely assume that $\langle x \rangle$ satisfies the property (1.11a). We show that $\langle x \rangle$ is convergent by using Lemma 1.7 and the following facts.

1.12. Fact. Any monoid homomorphism h: $N \to M$, M finite, factors as a composite

$$\begin{array}{ccc} N & \xrightarrow{\ e\ } & N_{k,l} \\ & \searrow^{h} & \downarrow{\scriptstyle i} \\ & & M \end{array} \tag{1.13}$$

where e is a surjective homomorphism, i is injective and $N_{k,l}$ was defined in the Preliminary section.

1.14. Fact. Any homomorphism h: $N \to N_{k,l}$ factors as

$$\begin{array}{ccc} N & \xrightarrow{\ k_n\ } & N_{n!,n!} \\ & \searrow^{h} & \downarrow{\scriptstyle h\#} \\ & & N_{k,l} \end{array} \tag{1.15}$$

where n is any number greater than both k and l. Note that the map h (and $h\#$) satisfies:

$$\begin{aligned} xh &= x \quad \text{if} \quad x < k; \\ &= k + \operatorname{Rem}(x-k, l) \quad \text{if} \quad x \geqq k. \end{aligned} \tag{1.16}$$

We now show that the property of (1.11a) implies that $\langle x\rangle$ is convergent. Suppose that $h\colon N\to M$ is a monoid homomorphism, M finite. Then by 1.12 and 1.14 we may as well assume that M is $N_{n!,n!}$ and h is defined by (1.16) when k and l are $n!$; thus we assume

$$\begin{aligned} zh &= z \quad \text{if} \quad z < n!; \\ &= n! + \operatorname{Rem}(z-n!, n!) \quad \text{if} \quad z \geqq n! \\ &= n! + \operatorname{Rem}(z, n!). \end{aligned}$$

It follows that $\langle xh\rangle$ is eventually constant, since $\langle x\rangle$ is strictly increasing, $x_m > n!$ for $m > n!$ and $x_m \equiv x_{m'}$ (mod $n!$), for sufficiently large m and m'. Thus, by (1.7), $\langle x\rangle$ is convergent. The proof is complete.

We now want to show that the monoid N^- is isomorphic to the (inverse) limit L of the diagram

$$\dots \to M_n \xrightarrow{g_n} M_{n-1} \xrightarrow{g_{n-1}} \dots \xrightarrow{g_2} M_1, \tag{1.17}$$

where $M_n = N_{n!,n!}$ and where the homomorphism $g_n\colon M_n \to M_{n-1}$ is defined, as one might expect by now, as

$$\begin{aligned} xg_n &= x \quad \text{if} \quad x < (n-1)!; \\ &= (n-1)! + \operatorname{Rem}(x, (n-1)!), \quad \text{if} \quad x \geqq (n-1)!. \end{aligned}$$

It is well known [G] that L may be described as the submonoid of πM_n consisting of all "compatible" sequences, i.e. all sequences $\langle z\rangle$, with $z_n \in M_n$, for each $n > 0$, and with $z_n g_n = z_{n-1}$, for $n > 1$. In order to prove N^- is isomorphic to L, we make two observations about the sequences in L.

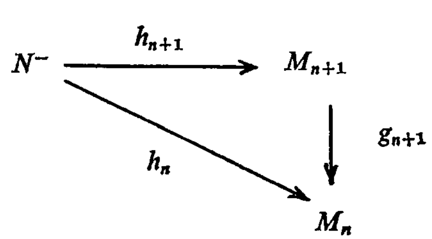
1.18. Lemma. If $\langle z\rangle$ is a compatible sequence and for some n, $z_n < n!$, then $z_m = z_n$, for all $m > n$.

1.19. Corollary. If $\langle z\rangle$ is a compatible sequence in L and $\langle z\rangle$ is not eventually constant, then for all $n > 0$,

$$n! \leqq z_n (< 2n!).$$

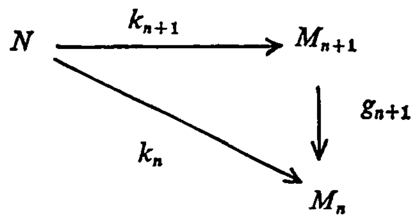
Now let $k_n\colon N \to M_n$ be the canonical homomorphism (i.e. $1k_n = 1$ in M_n) and let $h_n\colon N^- \to M_n$ be the monoid homomorphism taking the equivalence class of the sequence $\langle x\rangle$ to the eventual value of the sequence $\langle xk_n\rangle$ in M_n.

1.20. Lemma. For each $n > 0$ the diagram below commutes.



(1.21)

Proof. Since the diagram (1.22) commutes,



(1.22)

if a is the eventual value of $\langle xk_{n+1}\rangle$ then ag_{n+1} is also the eventual value of $\langle xk_n\rangle$.
Now since the monoid L, equipped with the projection maps

$$p_n\colon \langle z\rangle \to z_n$$

is the limit of the diagram (1.17), there is a unique homomorphism

$$q\colon N^- \to L$$

such that for each n, $q, p_n = h_n$. We want to show that q is an isomorphism.

1.24. Lemma. q is surjective.

Proof. Let $\langle x\rangle$ be a sequence in L. If $\langle z\rangle$ is eventually constant with value a then $\langle a\rangle^\sim q = \langle z\rangle$, where $\langle a\rangle^\sim$ is the equivalence class of the constant sequence $\langle a, a, \dots\rangle$. Otherwise $\langle z\rangle$ is strictly increasing. Let $\langle x\rangle$ be the sequence of integers with $x_n = z_n$ for all n (i.e. $\langle x\rangle$ is $\langle z\rangle$ considered as a sequence in N). Then we claim that for each n,

$$\langle x\rangle^\sim h_n = z_n$$

so that $\langle x\rangle^\sim q = \langle z\rangle$.
Indeed, for $m > n$, $x_m \geqq n!$ and

$$x_m \equiv z_n \pmod{n!}$$

by the definition of the g_m's. So the eventual value of $\langle xk_n\rangle$ is z_n. Thus q is surjective.

1.25. Lemma. q is injective.

Proof. Suppose that $\langle x\rangle^\sim q = \langle y\rangle^\sim q = \langle z\rangle$, where $\langle x\rangle$ and $\langle y\rangle$ are convergent sequences. We will show that $\langle x\rangle$ is equivalent to $\langle y\rangle$. Since this is clear in the case that $\langle x\rangle$ is eventually constant, assume that $\langle x\rangle$ is strictly increasing. Then the eventual value of the sequences $\langle xk_n\rangle$ and $\langle yk_n\rangle$ is z_n, so that, for all but finitely many values of m,

$$x_m \equiv y_m \pmod{n!}.$$

and hence $\langle x\rangle$ is equivalent to $\langle y\rangle$.
We have proved the following

1.26. Theorem. The monoid N^- is isomorphic to L.

It will sometimes be convenient to use L instead of N^- in order to get concrete representations. For example, notice how sequences in L are added. The n-th component of $\langle z\rangle + \langle z'\rangle$ is $z_n *_n z'_n$, where $*_n$ is the monoid operation in M_n. Thus, if both

$\langle z\rangle$ and $\langle z'\rangle$ are increasing, i.e. $n!\leqq z_n,\ z'_n$, then

$$z_n *_n z'_n = n! + \mathrm{Rem}\,(z_n + z'_n,\ n!).$$

Hence, aside from the notation "$n!$ +", the monoid operation *on increasing sequences in L* is the same as that on $\Pi(Z/n!:\ n>1)$, where $Z/n!$ is the ring of integers mod $n!$. Denoting the increasing sequences in L by **F**, we have seen that **F** forms a subring of the product of the rings $(Z/n!:\ n>1)$. (In fact, **F** is the inverse limit of the diagram

$$\dots \to Z/(n+1)! \to Z/n! \to \dots$$

where the maps are the canonical ring homomorphisms.) We turn now to the question of getting an explicit description of the sequences in L (and thus N^-). The description depends heavily on the following

1.27. "Factorial" lemma. For each $n\geqq 1$ and each number x, $0\leqq x<n!$, there is a *unique* sequence $a_1, \dots, a_{n-1}$ of integers with

$$0\leqq a_i\leqq i,\quad 1\leqq i<n$$

such that

$$x = \sum a_i i! = a_1 + a_2 2! + \dots + a_{n-1}(n-1)!.$$

(When $n=1$, the sequences is the empty sequence, whose sum is 0.)

This fact may be proved in a straightforward manner. Now, if $\langle z\rangle$ is an increasing sequence in L, for each n, $n!\leqq z_n<2n!$. Thus, we may write

$$z_n = n! + \sum_1^{n-1} a_i i!.$$

1.28. Proposition. If $\langle z\rangle$ is an increasing sequence in L, and for some n,

$$z_n = n! + \sum_1^{n-1} a_i i!, \quad \text{and}$$

$$z_{n+1} = (n+1)! + \sum_1^{n} b_i i!,$$

then for $i<n$, $b_i=a_i$.

Proof. Since $z_{n+1}g_{n+1}=z_n$, $\mathrm{Rem}\,(z_{n+1}, n!)=\sum a_i i!$ But $\mathrm{Rem}\,(z_{n+1}, n!)= =\sum b_i i!$. By the uniqueness part of the factorial lemma, the proposition is proved.

Hence, if $\langle z\rangle$ is an increasing sequence in L, there is a unique infinite *subdiagonal sequence* $\langle a\rangle$ (i.e. $0\leqq a_n\leqq n$, all n) such that

$$z_n = n! + \sum_1^{n-1} a_i i! \tag{1.28a}$$

for all n.

Let SD denote the set of all infinite subdiagonal sequences, and let SD^* denote the set of all finite subdiagonal sequences of integers. If $\langle a\rangle$ is any sequence in SD, let $z(a)$ denote the sequence defined by (1.28a) above. Then we have already proved part of the next theorem.

1.29. First representation theorem. For each sequence $\langle a\rangle$ in SD, $z(a)$ is an increasing sequence in L, and the map $SD\to \mathbf{F}$ defined by

$$\langle a\rangle \mapsto z(a)$$

is a bijection.

Proof. We already know the map is surjective. The fact that $z(a)$ is in fact in L (and hence in $\mathbf{F}$) is immediate, since $((n+1)!+\sum_1^n a_i i!)q_{n+1}=n!+\sum_1^{n-1} a_i i!$. If $\langle a\rangle$ and $\langle a'\rangle$ are distinct sequences in SD, then $z(a)$ and $z(a')$ are distinct by the factorial lemma. The theorem is proved.

It is easy to see that if $\langle z\rangle$ is an eventually constant sequence in L, then there is a unique finite subdiagonal sequence

$$\langle a_1, \dots a_{n-1}\rangle$$

such that for $k<n$

$$z_k = k!+\sum_1^{k-1} a_i i!,$$

and for $k>n$,

$$z_k = \sum_1^{n-1} a_i i!.$$

2. Topology

The standard topology on L (and thus on N^-) is inherited from the product topology on

$$\Pi M_n,$$

where each finite monoid M_n has the discrete topology. Thus the subbasis for the topology on L consists of all sets of the form

$$p_n^{-1}(X)$$

where p_n: $L\to M_n$ is the n-th projection map and X is some subset of M_n. Clearly L is Hausdorff in this topology, and is compact by the Tychonoff theorem since L is closed in the product. (Indeed, if $\langle z\rangle\notin L$, then for some n, $z_n g_n\neq z_{n-1}$. Then

$$p_n^{-1}(z_n)\cap p_{n-1}^{-1}(z_{n-1})$$

is an open set disjoint from L containing $\langle z\rangle$.) The subbasis sets are closed (as well as open) since

$$L-p_n^{-1}(X) = p_n^{-1}(M_n-X).$$

Thus L is a "Stone space". We will show that the Boolean algebra of the clopen (i.e. closed and open) subsets of L is isomorphic to the Boolean algebra of regular subsets of N.

Let i: $N\to L$ be the (unique) monoid homomorphism satisfying

$$i\cdot p_n = k_n \tag{2.0a}$$

for all $n>1$. The image of i consists of the eventually constant sequences. For $n>m$, let $g_{n,m}\colon M_n \to M_m$ be the composite of

$$M_n \xrightarrow{g_n} M_{n-1} \xrightarrow{g_{n-1}} \dots \xrightarrow{g_{m+1}} M_m$$

We will need the following fact.

2.1. Lemma. Suppose that $X \subset M_n$ and $Y \subset M_m$, where $n > m$. Then

$$\text{i)} \quad k_n^{-1}(X) \cap k_m^{-1}(Y) = k_n^{-1}(X \cap Z), \quad \text{and}$$

$$\text{ii)} \quad P_n^{-1}(X_n) \cap p_m^{-1}(Y) = p_n^{-1}(X \cap Z),$$

where $Z = g_{n,m}{}^{-1}(Y)$.

The easy proof is omitted.

2.2. Proposition. The subsets of L of the form

$$p_n^{-1}(X), \quad X \subset M_n,$$

are in fact a basis, being closed under finite union and finite intersection and complementation as well.

Proof. We have already noted the closure under complementation and closure under intersection follows from the preceding lemma, part ii). Closure under finite union follows from these two facts.

It follows from general topological principles that the clopen subsets of L are precisely the subsets of the form $p_n^{-1}(X)$, $n > 1$, $X \subset M_n$.

If S is a subset of N, let S^- denote the closure of $i(S)$ in L.

2.3. Lemma. Let $S = k_n^{-1}(X)$, where $X \subset M_n$. Then $S^- = p_n^{-1}(X)$. Thus the closure of a regular set is a clopen set in L.

Proof. Let $B = p_n^{-1}(X)$. Then S^- is contained in B since B is closed and $i(S) \subset \subset B$, by (2.0a). Let b be any point in B not in $i(S)$. We will show that b is a limit point of $i(S)$. Let $C = p_m^{-1}(Y)$ be any basis set containing b. We may assume that $n > m$. Applying 2.1,

$$B \cap C = p_n^{-1}(Z), \quad \text{and}$$

$$S \cap k_n^{-1}(Y) = k_n^{-1}(Z),$$

where $Z = g_{n,m}^{-1}(Y)$. Let R be the set $S \cap k_n^{-1}(Y)$. Then R is nonempty: otherwise, Z is empty, since k_n is surjective, and this would imply that $B \cap C$ would be empty, contradicting the fact that b is contained in this intersection. Now if $a \in R$, $i(a)$ is in $B \cap C$, by (2.0a) again, showing that b is a limit point of S. The proof of 2.3 is complete.

2.4. Lemma. Let R and S any subsets of N. Then $R \subset S$ iff $R^- \subset S^-$.

Proof. Only one direction is nontrivial. Suppose that R^- is contained in S^-. Let $x \in R$. Then the singleton set $\{x\}$ is a regular subset of N so the closure of $i(\{x\})$, say A, is a clopen subset of L, by 2.3 above. Since A is open and since $R^- \subset S^-$, A must contain a point of S. But since L is Hausdorff, A is a singleton set. Thus $x \in S$, proving the lemma.

2.5. Theorem. The map $R \mapsto R^-$ is a Boolean isomorphism from the Boolean algebra of regular subsets of N to the Boolean algebra of clopen subsets of L.

Proof. By Lemma 2.3, each regular subset of N is mapped to a clopen basis set and every clopen set in L is the closure of the image of a regular set. By Lemma 2.4, the map is an order (and hence Boolean) isomorphism.

The next fact follows from the standard proof of the Stone representation theorem.

2.6. Corollary. N^- is in bijective correspondence with the collection of ultrafilters on the Boolean algebra of regular subsets of N.

3. Some algebra

Recall that we have shown that the strictly increasing sequences **F** in L form a ring, isomorphic to the inverse limit of the diagram

$$\dots \to Z/n! \to Z/(n-1)! \to \dots$$

Although it is probably well known, we will indicate why **F** is also isomorphic as a ring to the product ring

$$\Pi(Z_p\colon\ p \text{ prime}),$$

where Z_p is the ring of p-adic integers.

One definition of Z_p (see [Kurosh, p. 154]) is the following:

3.1. Definition. Z_p consists of all sequences $\langle k\rangle$ of nonnegative integers satisfying

$$0 \leqq k_n < p^n, \tag{3.1a}$$

and

$$k_n \equiv k_{n+1} (\text{mod } p^n) \tag{3.1b}$$

for each $n>1$. The sequences are added and multiplied pointwise, where the n-th component is reduced modulo p^n. Thus Z_p is the inverse limit of the diagram

$$\dots \to Z/p^n \to Z/p^{n-1} \to \dots \tag{3.2}$$

Kurosh gives an easy proof that

3.3. Z_p has no zero divisors.

We will show that each sequence $\langle z\rangle$ in **F** determines a sequence $\langle z\rangle\alpha_p$ in Z_p.

3.4. Definition of α_p: $\mathbf{F} \to Z_p$.

Given $\langle z\rangle$, for each n let $m=p^n!$. Then for all $t>m$,

$$z_t \equiv z_m\ (\text{mod } m),$$

so that

$$z_t \equiv z_m\ (\text{mod } p^n).$$

Then for each n, we define k_n as $\text{Rem}\,(z_m, p^n)$, and we let $\langle z\rangle\alpha_p = \langle k\rangle$.

Taking the target tupling of the maps α_p, p prime, we get a map

$$\alpha\colon \mathbf{F} \to \Pi(Z_p\colon p \text{ prime}).$$

Since the ring operations on both $\mathbf{F}$ and Z_p are defined componentwise and since reduction mod m preserves addition and multiplication, we have

3.5. Lemma. Each map α_p is a ring homomorphism and thus so is α.

3.6. Lemma. α is an isomorphism.

Proof. First we show that α is injective. Let $\langle z\rangle$ and $\langle z'\rangle$ be distinct sequences in $\mathbf{F}$. Then for some n, $z_n \neq z'_n$. Factoring $n!$ as a product of powers of primes, there is some prime p and some m such that

$$z_n \not\equiv z'_n \,(\text{mod } p^m).$$

But for all $t > n!$,

$$z_t \equiv z_n \,(\text{mod } p^m)$$

and

$$z'_t \equiv z'_n \,(\text{mod } p^m).$$

Thus, $\langle z\rangle\alpha_p \neq \langle z'\rangle\alpha_p$, so that α is injective.

Now we prove α is in fact surjective. Suppose that we are given an element of $\Pi(Z_p\colon p \text{ prime})$, say

$$\langle y^p\rangle = \langle y^p(1), y^p(2), \ldots\rangle$$

for each prime p. We will construct a sequence $\langle z\rangle$ in $\mathbf{F}$ such that $\langle z\rangle\alpha_p = \langle y^p\rangle$ for each p. In order to do this, we construct an increasing sequence $\langle x\rangle$ in *Conv* with $x_1 = y^{p1}(1)$ (where $p1, p2, \ldots$, are the primes in increasing order) and for each $n > 1$,

$$x_n > x_{n-1}; \tag{3.7}$$

$$x_n \equiv y^{p1}(n)\,(\text{mod } (p1)^n)$$

$$x_n \equiv y^{p2}(n)\,(\text{mod } (p2)^n)$$

$$\ldots$$

$$x_n \equiv y^{pn}(n)\,(\text{mod } (pn)^n).$$

Suppose that $\langle x\rangle$ satisfies the above conditions. Then $\langle x\rangle$ is obviously increasing and is, although not so obviously, convergent. Indeed, to prove $\langle x\rangle$ is convergent we will show that for each n, all but finitely many of the terms of $\langle x\rangle$ are congruent modulo $n!$. Recall that if q and q' are relatively prime, then

$$u \equiv v \,(\text{mod } qq')$$

iff

$$u \equiv v \,(\text{mod } q) \quad \textit{and} \quad u \equiv v \,(\text{mod } q').$$

Now factor a fixed $n!$ into a product of primes, say

$$n! = p1^{k1} \ldots pr^{kr},$$

whre $0 \leqq ki$, $i = 1, \ldots, r$. Let m be any number greater than both r and $\max(k1, \ldots,$

... kr). Then for each $i<m$,

$$x_m \equiv y^{pi}(m) \pmod{pi^m},$$

so that

$$x_m \equiv y^{pi}(m) \pmod{pi^{ki}}$$

since $m>ki$. But since $\langle y^{pi}\rangle$ is in Z_{pi},

$$y^{pi}(m) \equiv y^{pi}(ki) \pmod{pi^{ki}}.$$

It follows that for $t>m$,

$$x_t \equiv y^{pi}(ki) \pmod{pi^{ki}}, \quad \text{all} \quad i<r.$$

But then, for $t>m$,

$$x_t \equiv x_m \pmod{n!},$$

proving that $\langle x\rangle$ is convergent. Let $\langle z\rangle$ be the sequence in L determined by $\langle x\rangle$. Then $\langle z\rangle\alpha_p=\langle y^p\rangle$, for all p.

It remains to show how to obtain the sequence $\langle x\rangle$. But assuming we have found $x_1, \dots, x_{n-1}$, we obtain x_n satisfying (3.7) by applying the Chinese Remainder Theorem.

The proof of the Theorem is complete.

4. Some closures

In section 2 we showed that if R is a regular subset of N, then the closure of $i(R)$ in L has the form

$$p_n^{-1}(X)$$

for some $X\subset M_n$.

4.1. Proposition. If R is an infinite regular subset of N, then R^- and thus its frontier are uncountable.

Proof. Suppose R is $k_n^{-1}(X)$, where $X\subset M_n$. If R is infinite, there is some element b in X with

$$n! \leqq b\,(<2n!).$$

Write b as $n!+\sum^{n-1} a_i i!$. Then, for all infinite subdiagonal sequences a which extend $\langle a_1, \dots, a_{n-1}\rangle$, the element $z(a)$ is in the closure of R. Since there are uncountably many such sequences a, the proposition is proved.

Now we prove that if S is the nonregular (context sensitive) set of squares, i.e.

$$S=\{n^2:\ n\in N\}$$

then the frontier of S is also uncountable.

We will use two facts from the preceding section.

(4.2) Z_p has no zero divisors.

(4.3) $\mathbf{F}$ is isomorphic as a ring to

$$\Pi(Z_p:\ p \text{ prime}).$$

Let $\langle x\rangle$ be any sequence in **F** and let $\langle y\rangle=\langle x\rangle\langle x\rangle=\langle x\rangle^2$ (so that for each n, $y_n=n!+\text{Rem}\,(x_n^2, n!)$).

4.4. Lemma. $\langle y\rangle$ is a limit point of S.

Proof. Let B be a basis set containing $\langle y\rangle$. Then $B=p_n^{-1}(X_n)$, for some n and some subset X_n of M_n. Thus $y_n\in X_n$. If a is the natural number

$$a = n!+x_n$$

then $k_n(a^2)=y_n$, so that $i(a^2)\in B$, proving that $\langle y\rangle$ is a limit point of S.

Now we will show that there are uncountably many points in L of the form $\langle x\rangle\langle x\rangle$, which will complete the proof that S^- is uncountable. Indeed, in any ring with no zero divisors, if $u^2=v^2$, then $u=v$ or $u=-v$, since

$$u^2-v^2 = (u+v)(u-v).$$

Now we choose in each ring Z_p two elements, say x_p and y_p, such that x_p is disctint from y_p and from $-y_p$.

Then, for each function f: Primes $\to\{0, 1\}$, let $z(f)$ be the element of $\Pi(Z_p$: p prime) defined by:

$$z(f)_p = x_p \quad \text{if} \quad f(p) = 0;$$
$$= y_p \quad \text{if} \quad f(p) = 1.$$

Then, by 4.4, regarding each element $z(f)$ as an element of **F**, $z(f)^2$ is a limit point of S. But if $f\neq f'$, then $z(f)^2\neq z(f')^2$; indeed, if $x_p=f(p)\neq f'(p)$, then $z(f)_p^2=x_p^2\neq y_p^2=z(f')_p^2$. We have proved:

4.5. Proposition. S has an uncountable frontier.

Thus the cardinality of the frontier of a language cannot distinguish regular from nonregular sets. In the next section we will characterize all sets whose frontier is uncountable.

5. Explicit topology

In this section we will use the first representation theorem to get a second, more geometric description of L and the topology on L. This description makes use of a locally finite rooted tree T (see [EBT]). The vertices of T are certain finite sequences in $SD^*\cup SD^*\times\{\perp\}$.

A vertex of T in the form of a pair $(\langle s\rangle, \perp)$ is written as a finite sequence $\langle s, \perp\rangle$ ending in the symbol $\perp$. The root of T is $\langle\perp\rangle$; the root has three immediate successors: $\langle\rangle$ (the empty sequence), $\langle 1\rangle$ and $\langle 0, \perp\rangle$; if

$$v = \langle a_1, \dots, a_n, \perp\rangle$$

is a vertex in T, then $\langle a_1, \dots, a_n\rangle$ is in SD^*, and v has the following $2n+3$ immediate successors:

$$\langle a_1, \dots, a_n, 1\rangle, \dots \langle a_1, \dots, a_n, n+1\rangle \quad \text{and}$$
$$\langle a_1, \dots, a_n, 0, \perp\rangle, \dots, \langle a_1, \dots, a_n, n+1, \perp\rangle.$$

All vertices in T not ending in $\perp$ are leaves; all vertices ending in $\perp$ are not.

A (root) *path* in T is a sequence $v_0=\bot, v_1, \ldots$ of vertices, perhaps infinite, such that for each n, v_{n+1} is an immediate successor of v_n. We let P denote the collection of all root paths which are either infinite or are finite and end in a leaf.

We want to prove that N^- is in bijective correspondence with P. In order to see this, we define two maps

$$\text{val}: \text{Ver} \to N$$

$$\text{path}: N \to P$$

where Ver is the set of vertices of T, as follows:

$$\text{val}\,\langle a_1, \ldots, a_{n-1}\rangle = \sum_1^{n-1} a_i i!;$$

$$\text{val}\,\langle a_1, \ldots, a_{n-1}, \bot\rangle = n!+\sum_1^{n-1} a_i i!.$$

Note that val $\langle\rangle=0$, where $\langle\rangle$ is the empty sequence.

Path is defined as follows: for each x in N, path (x) is the root path in P to the leaf vertex v, where val $v=x$. Note that if $x>0$, then $a_{n-1}>0$, and n is least number such that $x<n!$. (Of course, path$(0)=\langle\rangle$.) Lastly, if $v=\langle a_1, \ldots, a_{n-1}, \bot\rangle$, let $l(v)=n!$.

We note one important fact.

Fact. If there is a path from v to v' in T, then

$$\text{val}\,(v) \equiv \text{val}\,(v') \pmod{l(v)}.$$

We now define a function from L to P.

Definition of p: $L\to P$.

If $\langle z\rangle$ is an eventually constant sequence with value x in N, then $\langle z\rangle p=\text{path}\,(x)$; if $\langle z\rangle$ is a strictly increasing sequence in L, then $\langle z\rangle p$ is the infinite root path

$$\bot,\ v_1, \ldots, v_n, \ldots$$

where

$$v_n = \langle a_1, \ldots, a_{n-1}, \bot\rangle$$

$$\text{if}\quad z_n = n!+\Sigma a_i i!,\quad \text{for}\quad n>1.$$

Now we equip the set P of maximal root paths in the tree T with a topology as follows. For each vertex v of T, let $B(v)$ denote the set of all paths in P which contain v.

5.1. Definition. The topology on P is determined by taking the collection of sets $B(v)$, $v\in\text{Ver}$, as a basis.

Note that if v and v' are incomparable vertices, then $B(v)\cap B(v')$ is empty, and if $v<v'$, the intersection is $B(v')$.

Recall the bijection p: $L\to P$ above. The topology on L can be easily "seen" in P.

5.2. Theorem. p is a homeomorphism.

Proof. Let $B=B(v)$ be a basis set in P. If v is a leaf, $B(v)$ is a singleton and $p^{-1}(B)$ consists of the eventually constant sequence with eventual value val$\,(v)$.

Otherwise,

$$v = \langle a_1, \ldots, a_{n-1}, \perp\rangle \text{ so that}$$

$$p^{-1}(B) = \{\langle z\rangle \in L\colon\ z_n = n! + \Sigma a_i i!\} = p_n^{-1}(\text{val } v).$$

In either case, $p^{-1}(B)$ is a basis set in L. Thus, p is continuous. The argument that p^{-1} is also continuous is equally easy and is omitted.

The *frontier* of a subset A of a topological space is $A^- - A$. Using the above geometric picture of the topology on P, we may describe the frontiers of subsets A of N as follows. Each element x of A determines a finite path path (x) in P (ending in a leaf v with val $(v)=x$.) The collection of all the vertices in path (x) for $x \in A$, determines a subtree of T, say $T(A)$. The important fact about the tree representation is this: if we identify the elements of N with their images under path, we obtain

The second representation theorem. *The infinite paths in $T(A)$ are precisely the elements in the frontier of A.*

For example, if we want to find a set A whose frontier is only countably infinite, we might want $T(A)$ to "look like" the tree in figure 5.2a. To do this, one may define A as the set of all numbers of the form val $\langle 1, 2, \ldots, n, 0, \ldots, 0, 1\rangle$ (for $n, m > 0$, where there are m 0's).

Since we want to characterize those subsets of N whose frontier in L (or P or N^-) is uncountable, we will prove a theorem concerning those locally finite trees that have an uncountable number of infinite paths.

5.3. Definition. B_2 is the complete binary tree — i.e. each vertex in B_2 has exactly two immediate successors.

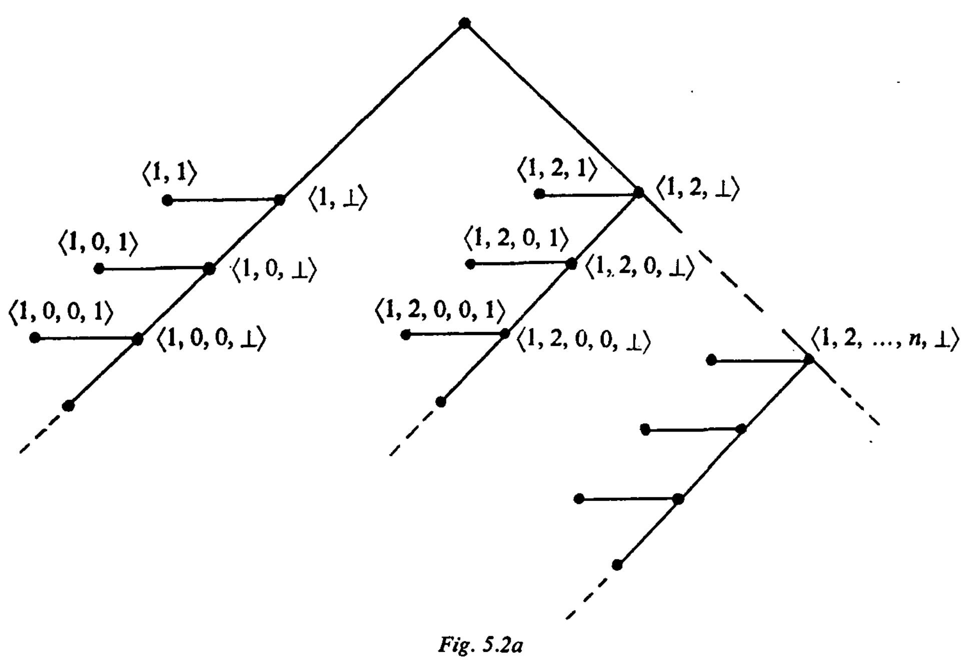


Fig. 5.2a

5.4. Definition. Let $T_i=(V_i, E_i)$, $i=1,2$ be rooted trees, with V_i the set of vertices and E_i, the set of (ordered) edges. An *order embedding* $T_1 \to T_2$ is function $f\colon V_1 \to V_2$ such that for each pair $\langle v, v' \rangle$ of vertices in T_1, there is a path in T_1 from v to v' iff there is a path from vf to $v'f$ in T_2.

5.5. Theorem. Let T be a rooted locally finite tree. Then T has an uncountable number of infinite paths iff there is an order embedding of $B_2 \to T$.

Proof. Since clearly B_2 has uncountably many paths, it is easy to see that if there is an order embedding of B_2 in T, T has uncountably many paths as well. Now to prove the converse, we use the following fact

5.6. Lemma. Let T be a locally finite tree with uncountably many paths. Then there are two incomparable vertices v_1 and v_2 in T (i.e. it is not the case that $v_1 < v_2$ or $v_2 < v_1$) such that for $i=1,2$, $T(v_i)$ has uncountably many paths, where $T(v_i)$ is the subtree of T consisting of v_i and all of its successors.

Proof of the Lemma. Since T is locally finite, for each n there are only finitely many vertices in T of depth n. For some n there must be two distinct vertices at depth n, say v_1 and v_2, such that the subtrees $T(v_1)$ and $T(v_2)$ of all descendents of v_1 and v_2 respectively both have uncountably many paths. Otherwise, T has only countably many paths, a contradiction.

Using this lemma, we can define an order embedding of B_2 in T, by induction on the depths of the vertices in B_2. The root of B_2 maps to the root of T. The two successors of the root of B_2 map to the first pair of incomparable vertices v_1 and v_2 in T such that $T(v_i)$, $i=1,2$ has uncountably many paths. Having defined the embedding f on all vertices of B_2 of depth n such that for a vertex v in B_2, the tree $T(vf)$ has uncountably many paths, we use the lemma again to extend the definition of f one level further. The proof of the theorem is complete.

We may easily translate this result into an arithmetic form. For integers u and v, define

$$u \subseteqq v \quad \text{if} \quad u \leqq v \quad \text{and} \quad u \equiv v \pmod{l(u)};$$

i.e. if

$$u = n! + \sum a_i i!$$

and

$$v = m! + \sum b_j j!,$$

then $u \subseteqq v$ iff $n \leqq m$ and $a_i = b_i$ for $i<n$. Say that a subset W of N is "B_2-like" if for all x in W there are y, z in W such that $x \subseteqq y$ and $x \subseteqq z$ and y and z are $\subseteqq$-incomparable.

The translation of 5.5 is:

5.7. Proposition. A subset X of N has an uncountable frontier iff X contains a nonempty B_2-like subset.

We obtain the following number theoretic fact as a result.

5.8. Corollary. The set of squares contains a B_2-like subset.

6. Final Remarks

It appears that the cardinality of the frontier of a one letter language is not a useful tool for making distinctions among languages. One might then ask whether the consideration of the sequence of frontiers of a language $X \subset N$ will be more useful, where the sequence $X = X_0, X_1, \dots$ is defined by

$$X_{n+1} = X_n^- - X_n;$$

i.e. the $n+1$-st set is the frontier of the n-th. However, it is easy to show that for every subset X of N, X_2, and hence X_n for $n > 2$, is empty.

In Scott's talk, several of the results given here were stated: the theorem in Section 3 that **F** is isomorphic to the product of the p-adic integers; the fact that N^- formed a compact, zero-dimensional Hausdorff space and our Corollary 2.6; most importantly he stated a version of the first representation theorem for the sequences in **F**.

The paper [B] has some results of a category-theoretic nature related to the theorem in Section 3. We have not made any use of the fact that N^- forms a free profinite monoid. The reader interested in other properties of N^- (and other free profinite monoids) may consult [B2] and [R].

Some of the results in Section 1 and 2 can be generalized to the case of the structure of convergent sequences of words in an arbitrary alphabet. The interesting problem of finding a concrete representation of the equivalence classes of these sequences is, as far as I know, still open. However, in the case that M is a finitely generated free *commutative* monoid, the monoid M^- is a finite powei of N^-, as Z. Esik observed.

Addendum

In a recent conversation, Scott suggested modifying the definition of the frontier of a subset S of N as follows. Instead of defining the frontier of S as the closure of S^- minus S, $S^- - S$, let:

$$\text{fron}\,(S) = S^- - \text{int}\,(S^-),$$

where "int" denotes the interior operator. This definition has the property that exactly the regular sets have an empty frontier, since $\text{fron}(S) = \emptyset$ iff the closure of S is clopen.

Does the cardinality of this "new frontier" give more information about the structure of the set? Not much. For example, when S is the set of squares, we have already shown that the closure of S is uncountable. The interior of the closure is empty, by the following observation.

Proposition.. Let S be an infinite subset of N. If C is the closure of S, $cl(S)$, $O = \text{int}\,(C) \setminus F$ is empty iff S contains no infinite regular subset. (Recall that F is the set of infinite elements of $cl(N)$, the closure of N.)

Proof. First suppose that S contains no infinite regular subset. If $x \in O$, there is a regular subset R of S such that $x \in cl(R)$, by the definition of the topology. Since R is finite, $cl(R) = R$, and x itself is a finite number. Now suppose that $\text{int}(C) \setminus F$ is empty but that R is an infinite regular subset of S. Then $cl(R)$ is a clopen subset of $\text{int}(C)$ and contains uncountably many elements of F.

Applying this proposition to the set Sq of squares, we see that $\text{int}\,(cl(Sq))$ contains only the elements of Sq itself, since, by the pumping lemma, Sq contains no infinite regular subset. Thus, the cardinality of $\text{fron}(Sq)$ is uncountable.

We now show how to construct, for any subset A of integers, a set $S=S(A)$ such that fron(S) is also uncountable, and such that A is Turing equivalent to S. (Thus, if A is nonrecursive, so is S.) In this construction, we make use of the representation of the elements of $cl(N)$ as paths in the tree T defined in section 5, as well as proposition 5.7. We define the vertices in a subtree $A(T)$ of T by induction: First assume that the vertex $\langle 0, 0, 0, \perp\rangle$ belongs to $A(T)$;
Now assume that for $4<n$, the set of vertices V_{n-1} of length $<n$ which belong to $A(T)$ have been defined. For each vertex $v \in V_{n-1}$ which is not a leaf, write v as

$$v = \langle a_1, \ldots, a_{n-1}, \perp\rangle.$$

Then define:
$v1$ $(=\langle a_1, \ldots, a_{n-1}, 1\rangle) \in V_n$ iff $n \in A$; furthermore, define, for $j=0$ and $j=1$
$vj\perp$ $(=\langle a_1, \ldots, a_{n-1}, j, \perp\rangle) \in V_n$ iff $n \in A$; if n is not in A, then for $j=2$ and 3, $vj\perp \in V_n$.

Otherwise, vj and $vj\perp$ do not belong to V_n.
This completes the definition of the set of vertices of the tree $A(T)$.

Note that the tree B_2 may be order embedded in $T(A)$, so that by proposition 5.7 the set S determined by the leaves of $T(A)$ has an uncountable closure. Moreover, $T(A)$ does not contain all infinite paths containing a particular vertex, and hence S does not contain any infinite regular subset. Lastly, it is clear that A and S are Turing equivalent.

Acknowledgements

It is a pleasure to thank Dana Scott for introducing me to his convergent sequences and sending me some references, including the slides of this talk. Jess Wright, Ralph Tindell and Doug Troeger made several useful suggestions.

Abstract

In a talk titled "Infinite words" given in the spring of '83, Dana Scott introduced the notion of a convergent sequence of "words" (i.e. elements of a finitely generated free monoid). Scott stated a number of properties of the collection of (equivalence classes of) convergent sequences, some of which showed that this structure forms a free "profinite" monoid. The monoid of convergent sequences has a natural Stone space topology in which subsets of the free monoid have closures. Scott asked whether an investigation of the "frontiers" of languages would lead to a useful classification of languages. In this paper, an explicit description of the frontiers of all subsets of N, the one-generated free monoid, is obtained. It is shown that the cardinality of the frontier of a subset of N cannot distinguish regular from nonregular sets. A necessary and sufficient condition that a subset of N have an uncountable frontier is given.

DEPARTMENT OF PURE AND APPLIED MATHEMATICS
STEVENS INSTITUTE OF TECHNOLOGY
HOBOKEN, NJ 07030

CONSULTANT TO THE MATHEMATICAL SCIENCES DEPARTMENT
IBM WATSON RESEARCH CENTER
YORKTOWN HEIGHTS. NY 10598

References

[B] BANASCHEWSKI, B., On profinite universal algebras. *General topology and its relations to modern analysis and algebra II* (Preceedings 3rd Prague Topology Symposium 1971). Academia Prague 1972, 51—62.
[B2] BANASCHEWSKI, B., The Birkhoff Theorem for Varieties of Finite Algebras, manuscript 1980.
[EBT] ELGOT, C., BLOOM, S. L., TINDELL, R., The algebraic structure of rooted trees, *J. of Computer and System Science*, vol. 16, No. 3 (1978) 362—399.
[G] GRATZER, G., *Universal Algebra* (Van Nostrand, Princeton, Toronto, London; 1968).
[K] KUROSH, *The theory of groups*, translated by K. Hirsch, (Chesea, New York; 1960)
[R] REITERMAN, J., The Birkhoff theorem for finite algebras; *Algebra Universalis* 14 (1982) 1—10.
[S] SCOTT, D., Infinite words, slides of a lecture given during the spring of 1983.

(Received Febr. 22, 1984)