SIAM J. ALG. DISC. METH.
Vol. 3, No. 3, September 1982

0196-5212/82/0303-0011 $01.00/0

MULTISETS OF APERIODIC CYCLES*

N. G. DE BRUIJN† AND D. A. KLARNER‡

Abstract. The basic result is that if A is a finite set then there are exactly $|A|^n$ multisets of aperiodic cycles over A with total length n. This is shown by a counting technique but also by establishing an explicit bijection from these multisets to words over A of length n.

1. Notation. Let $\mathbb{N}=\{0, 1, \cdots\}$ and $\mathbb{P}=\{1, 2, \cdots\}$ be the sets of nonnegative and positive integers respectively. Let A be a finite or countable set, and for each $n \in \mathbb{P}$ let A^n be the set of all n-tuples over A. An element $x \in A^n$ is called an *n-word* or sometimes just a *word*, and the *length* of x is defined to be $\lambda(x)=n$. Also, it is convenient to have an *empty word* Λ whose length is defined to be $\lambda(\Lambda)=0$. Let A^* be the set of all words over A, then elements $x, y \in A^*$ are concatenated to form a new word $xy \in A^*$ in the usual way. In particular, $\Lambda x = x\Lambda = x$ for all $x \in A^*$. Also, for each $k \in \mathbb{P}$, $x \in A^*$, x^k is defined to be the concatenation of k copies of x. (That is, $x^1 = x$, and $x^{k+1} = x \cdot x^k$ for all $k \in \mathbb{P}$). If $x, y \in A^*$ are such that $y = xu$ for some $u \in A^*$, we write $x \subseteq y$ and say x is an *initial word* of y.

Suppose $x \in A^n$, $n \in \mathbb{P}$, with $x = x_1 \cdots x_n$, $x_i \in A$ for $i = 1, \cdots, n$. Then the word $x_{d+1} \cdots x_n x_1 \cdots x_d$ is called the *d-shift* of x for $d = 1, \cdots, n-1$. If x is equal to a d-shift of x, then x is called *periodic*; otherwise, x is called *aperiodic*. The empty word is neither periodic, nor aperiodic. Suppose x is periodic and let $d \in \mathbb{P}$ be the smallest number such that x is equal to the d-shift of x. Then d is called the *period* of x and we define $\pi(x)$ to be the initial word of x with length d. If x is aperiodic we define $\pi(x) = x$. It is easy to check that $\pi(x)$ is an aperiodic word for all $x \in A^* \backslash \{\Lambda\}$. Also, if x is an n-word and $\pi(x)$ is a d-word where $n \in \mathbb{P}$, then d divides n. Furthermore, if $k = n/d$, then $x = (\pi(x))^k$. This motivates the definition of a *k-fold word*, namely, a word having the form y^k with y an aperiodic word. Let $P_k(A)$ be the set of k-fold words over A; in particular, $P_1(A)$ is the set of aperiodic words.

Now we define an equivalence relation on $A^* \backslash \{\Lambda\}$. Put $x \sim y$ just when $x = y$ or y is a d-shift of x for $1 \leqq d < \lambda(x)$. An equivalence class is called a *cycle* over A, and $\mathscr{C}(A)$ is defined to be the set of all cycles over A. If $x \in A^* \backslash \{\Lambda\}$, let $\langle x \rangle$ be the cycle which has x as an element. We will speak of a property common to all the words in an equivalence class as a property of the cycle, and notation will be used in a similar fashion even though this is a bit improper. For example, if $x \sim y$, then $\lambda(x) = \lambda(y)$. So we speak of the length of $\langle x \rangle$ and write $\lambda\langle x \rangle$ for it. Also, if $x \sim y$, then x and y have the same period, and $\pi(x) \sim \pi(y)$. So we speak of the period of $\langle x \rangle$, and write $\pi\langle x \rangle = \langle \pi(x) \rangle$. If x is a k-fold word, $\langle x \rangle$ is called a *k-fold cycle*. Let $\mathscr{A}_k(A)$ be the set of k-fold cycles over A, $k \in \mathbb{P}$. If x is aperiodic, we say $\langle x \rangle$ is aperiodic. Let $\mathscr{A}(A) = \mathscr{A}_1(A)$ be the set of aperiodic cycles over A. Thus, $\{\mathscr{A}_1(A), \mathscr{A}_2(A), \cdots\}$ is a partition of $\mathscr{C}(A)$ into disjoint sets.

A *finite multiset* on a set X is a mapping f from X into $\mathbb{N}$ such that the *size* of f, which is defined to be $\omega(f) = \sum_{x \in X} f(x)$, is finite. In this paper we shall just say *multiset* instead of "finite multiset". Let $\mathscr{M}(A)$ and $\mathscr{M}_k(A)$ be the set of all multisets on $\mathscr{C}(A)$ and $\mathscr{A}_k(A)$ respectively for all $k \in \mathbb{P}$. For f in $\mathscr{M}(A)$ or in $\mathscr{M}_k(A)$ we introduce

* Received by the editors July 14, 1981, and in revised from November 19, 1981.

† Eindhoven University of Technology, Department of Mathematics, 5600 MB Eindhoven, the Netherlands.

‡ Mathematical Sciences Department, State University of New York, Binghamton, New York, 13901.

a measure $\kappa(f)$ called the length total of f, that is different from the size $w(f)$. Instead of just counting the elements of the multiset, we give each element C a weight that equals $\lambda(C)$. So the *length total* of $f \in \mathcal{M}(A)$ is defined to be $\kappa(f) = \sum_{C \in \mathcal{C}(A)} f(C)\lambda(C)$, and $\kappa(f)$ for $f \in \mathcal{M}_k(A)$ is defined similarly as a sum over $\mathcal{A}_k(A)$. If $\kappa(f) = n$, we call f an *n-multiset* for all $n \in \mathbb{N}$.

2. Statement of results. The heart of our results is a bijection between the set of n-multisets over $\mathcal{A}(A)$ and the set of n-words over A. That is, every multiset of aperiodic cycles whose lengths total n corresponds to a word of length n, and conversely. We will use this bijection to prove a weighted version. Let w be a product weight on A^*. That is, w is a mapping of A^* into some sort of commutative algebra, and one of the most important properties of w is that $w(xy) = w(x)w(y)$ for all $x, y \in A^*$. This means $w(x_1 \cdots x_n) = w(x_1) \cdots w(x_n)$ for all $x_i \in A$, $i = 1, \cdots, n$. For our purposes, w must possess some other properties which guarantee that certain infinite sums and products are themselves weights. These extra assumptions will become evident later, but will not be explicitly stated. Since $x \sim y$ implies $w(x) = w(y)$, we speak of the weight of a cycle $C \in \mathcal{C}(A)$ and write $w(C) = w(x)$ for all $x \in C$.

Finally we define a weight function W on $\mathcal{M}(A)$ and $\mathcal{M}_k(A)$ in terms of w. For all $f \in \mathcal{M}(A)$ we put

$$W(f) = \prod_{C \in \mathcal{C}(A)} (w(C))^{f(C)}. \tag{1}$$

The weight of $f \in \mathcal{M}_k(A)$ is defined similarly as a product over $\mathcal{A}_k(A)$, $k \in \mathbb{P}$. Note that since $\{\mathcal{A}_1(A), \mathcal{A}_2(A), \cdots\}$ is a partition of $\mathcal{C}(A)$, a multiset $f \in \mathcal{M}(A)$ is completely determined by its restrictions to the blocks of this partition. Let f_k be the restriction of f to $\mathcal{A}_k(A)$ for all $k \in \mathbb{P}$. Then it is easy to see that

$$W(f) = \prod_{k=1}^{\infty} W(f_k). \tag{2}$$

and furthermore,

$$\sum_{f \in \mathcal{M}(A)} W(f) = \prod_{k=1}^{\infty} \sum_{f_k \in \mathcal{M}_k(A)} W(f_k). \tag{3}$$

Finally, we will prove that for all $k \in \mathbb{P}$

$$\sum_{f \in \mathcal{M}_k(A)} W(f) = \sum_{g \in \mathcal{M}_1(A)} (W(g))^k. \tag{4}$$

Here we are dealing with multisets of k-fold cycles. Recall that there is a natural bijection π between $\mathcal{A}_k(A)$ and $\mathcal{A}_1(A)$. Say $\langle x \rangle \in \mathcal{A}_k(A)$, then $\langle \pi(x) \rangle$ is the corresponding element in $\mathcal{A}_1(A)$. Going the other way, $\langle y \rangle \in \mathcal{A}_1(A)$ corresponds to $\langle y^k \rangle$ in $\mathcal{A}_k(A)$. Thus, $g \in \mathcal{M}_1(A)$ corresponds to $f \in \mathcal{M}_k(A)$ where $g\langle x \rangle = f\langle x^k \rangle$ for all $\langle x \rangle \in \mathcal{A}_1(A)$; also, $W(f) = (W(g))^k$ for all $k \in \mathbb{P}$.

We will prove in subsequent sections that

$$\sum_{f \in \mathcal{M}_1(A)} W(f) = \sum_{x \in A^*} w(x) = \sum_{n=0}^{\infty} \left(\sum_{a \in A} w(a) \right)^n. \tag{5}$$

However, an immediate consequence of (5) is that

$$\sum_{f \in \mathcal{M}_1(A)} (W(f))^k = \sum_{n=0}^{\infty} \left(\sum_{a \in A} (w(a))^k \right)^n. \tag{6}$$

This follows just by replacing w with the kth power of w in the definition of the weight of $f \in \mathcal{M}_1(A)$ so that W is replaced with the kth power of W. Combining (3), (4), and (6) gives

$$\sum_{f \in \mathcal{M}(A)} W(f) = \prod_{k=1}^{\infty} \sum_{n=0}^{\infty} \Big(\sum_{a \in A} (w(a))^k \Big)^n. \tag{7}$$

However, we still have to prove (5); this is done combinatorially in § 4 and algebraically in § 5.

3. Some motivation. The inspiration for this paper was a formula due to Read [8] which expresses $f(n)$, the number of endomorphism patterns on an n-set, in terms of $t(1), \cdots, t(n)$, where $t(k)$ is the number of isomorphism classes of rooted trees with n points. The formula is most elegantly expressed in terms of the two generating functions involved. Let $F(z) = 1 + f(1)z + f(2)z^2 + \cdots$, and let $T(z) = t(1)z + t(2)z^2 + \cdots$. Then Read's formula is

$$F(z) = \prod_{k=1}^{\infty} \frac{1}{1 - T(z^k)}. \tag{8}$$

Read's derivation of (8) was based on a formula due to Harary [6] who used Pólya's fundamental enumeration theorem to find an algorithm for computing the number of endomorphism patterns with n vertices.

An endomorphism on a set of n points can be described by means of a collection of cycles where each point of each cycle is the root of a tree. The idea that gave rise to the present paper is the discovery that the separate factors in (1) allow simple interpretations. The factor $(1 - T(z))^{-1}$ is related to the cycles on which the set of trees does not show any periodicity, and in general the factor $(1 - T(z^k))^{-1}$ is related to the cycles on which the set of trees has the exact period m/k (where m is the number of points on the cycle). A description of how endomorphisms are related to cycles with trees is given in § 6.

A quite simple observation is that the counting argument (presented in its simplest form in § 5) does not make use of the fact that the objects growing on the cycles are trees. We can replace the trees by the elements of any arbitrary finite or countable set A. This gives rise to the problem formulation presented in § 2.

4. The bijection. Our objective in this section is to construct a bijection Ω between the set of n-words over A and the set of n-multisets of aperiodic cycles over A represented in a certain normal form. Also, Ω preserves weights. That is, if x corresponds to multiset f, then $w(x) = W(f)$. Since

$$\sum_{x \in A^n} w(x) = \Big(\sum_{a \in A} w(a) \Big)^n,$$

we can sum this over all $n \in \mathbb{N}$ and use the bijection Ω to get (5).

The basis of Ω is an algorithm which factors an n-word into its corresponding n-multiset expressed in normal form. The algorithm and normal form both depend on an arbitrary linear order imposed on A. So we suppose the countable set A is ordered linearly by $\leqq$, and extend $\leqq$ to the lexicographical order on A^* in the usual way. That is, for all $x, y \in A^*$, $x \leqq y$ means $x \subseteq y$ or there exist $u, r, s \in A^*$ and $p, q \in A$ with $p < q$ such that $x = upr$, $y = uqs$.

We should not be trapped into thinking that for $u, v, x \in A^*$, the inequality $u < v$ always implies $ux < vx$. The implication is correct, however, if $\lambda(u) = \lambda(v)$. We shall use this repeatedly.

A word $x \in A^*$ is called *normal* if $\lambda(x) = 1$, or if $\lambda(x) > 1$ and x is less than all its d-shifts for $d = 1, \cdots, \lambda(x) - 1$. If x is aperiodic, then x and all its d-shifts ($d = 1, \cdots, \lambda(x) - 1$) are distinct.

Let $N(A)$ be the set of normal words over A. Thus, every aperiodic cycle contains exactly one normal word, and every normal word x gives rise to an aperiodic cycle x. The sequence of normal words $(c_1, \cdots, c_k)$ is defined to be the *normal form* $\nu(f)$ of a multiset f over $\mathcal{A}(A)$ just when $c_1 \geqq \cdots \geqq c_k$, $\kappa(f) = \lambda(c_1) + \cdots + \lambda(c_k)$, and for each $c \in N(A)$ the number of i ($1 \leqq i \leqq k$) with $c_i = c$ equals $f(c)$. Roughly speaking, normal words replace aperiodic cycles, and f is represented as a decreasing list of words after this replacement. For example, suppose $A = \{a, b, c, \cdots\}$ with $a < b < c < \cdots$, and define f to be 0 for all aperiodic cycles over A except that $f\langle a\rangle = 3$, $f\langle b\rangle = 1$, $f\langle ab\rangle = 2$, and $f\langle abacb\rangle = 1$. This means $\kappa(f) = 3 + 1 + 4 + 5 = 13$, and since b, $abacb$, ab, a are all normal and in decreasing order, $\nu(f) = (b, abacb, ab, ab, a, a, a)$. It will turn out that the word corresponding to f in this case is $\Omega^{-1}(f) = babacbababaaa$.

Let $x \in A^*$ with $x \neq \Lambda$, let $\sigma(x)$ be the longest normal initial word of x, and let $\tau(x)$ be the rest of x after $\sigma(x)$ has been deleted. Create a sequence $\Omega(x)$ of normal words as follows. Define $\Omega(x) = (x)$ for all normal words x, and define $\Omega(x) = (\sigma(x), \Omega(\tau(x)))$ for all other nonempty words. (Delete extra parentheses according to the rules $(u, (v)) = (u, v)$.)

If $\Omega(x) = (x_1, \cdots, x_n)$ then $x = x_1 \cdots x_n$. We shall show in Lemmas 3 and 4 that the converse is true if $x_1, \cdots, x_n$ are all normal and $x_1 \geqq \cdots \geqq x_n$.

LEMMA 1. *Suppose $x, y \in A^*$, y is normal, $x \neq \Lambda$ and $x < y$. Then $xy < yx$.*

Proof. By definition of $x < y$, there are two cases. In the first case, we have $y = xt$ with $t \in A^*$, $t \neq \Lambda$. Then $xt < tx$ because y is normal, so $xy = xxt < xtx = yx$. In the second case, we have $x = upr$, $y = uqs$ with $u, r, s \in A^*$, $p, q \in A$, and $p < q$. Then $xy = upry < uqsx = yx$ because $p < q$. This completes the proof. □

The next result is a generalization of this one.

LEMMA 2. *Suppose $x, y_1, \cdots, y_k \in A^*$, $y_1, \cdots, y_k$ normal, and $x \leqq y_1, \cdots, y_k$. Then $xy_1 \cdots y_k \leqq y_1 \cdots y_k x$.*

Proof. Using the previous lemma we have

$$xy_1 \cdots y_k \leqq y_1 xy_2 \cdots y_k \leqq y_1 y_2 x \cdots y_k \leqq \cdots \leqq y_1 y_2 \cdots y_k x. \tag{9}$$

That is, the ith inequality holds because $x \leqq y_i$ and because y_i normal implies $xy_i \leqq y_i x$ for $i = 1, \cdots, k$. This completes the proof. □

LEMMA 3. *Suppose $x_1, \cdots, x_h \in A^*$, $x_1, \cdots, x_h$ are normal, and $x_1 \geqq \cdots \geqq x_h$. Let $x = x_1 \cdots x_h$. Then $\Omega(x) = (x_1, \cdots, x_h)$. That is, every n-multiset over $\mathcal{A}(A)$ in normal form is the image under Ω of some element of A^n.*

Proof. It is enough to show that $\sigma(x) = x_1$, because then $\Omega(x) = (x_1, \cdots, x_n)$ follows by a simple induction. Since x_1 is normal, we certainly have $x_1 \subseteq \sigma(x)$. Suppose $\sigma(x) = x_1 \cdots x_k u$ where $u \subseteq x_{k+1}$ for some $k = 1, \cdots, h-1$, and $u \neq \Lambda$. Then $x_1, \cdots, x_k \geqq x_{k+1} \geqq x$, and $x_1, \cdots, x_k$ are normal by hypothesis, so $ux_1 \cdots x_k \leqq x_1 \cdots x_k u$ by Lemma 2. This means $x_1 \cdots x_k u$ is not normal. Thus, $\sigma(u) \subseteq x_1$, so $\sigma(u) = x_1$. This completes the proof. □

LEMMA 4. *Suppose $x \in A^*$, $x \neq \Lambda$, $\Omega(x) = (x_1, \cdots, x_k)$. Then $x_1 \geqq \cdots \geqq x_k$.*

Proof. It can be assumed without loss of generality that $k = 2$. If $k = 1$, there is nothing to prove. If the theorem is true for $k = 2$, then for $k \geqq 3$ one can use the fact that $\Omega(x) = (x_1, \cdots, x_k)$ implies $\Omega(x_1 x_{i+1}) = (x_i, x_{i+1})$ for $i = 1, \cdots, k-1$ to conclude that $x_i \geqq x_{i+1}$ for $i = 1, \cdots, k-1$; that is, $x_1 \geqq \cdots \geqq x_k$.

We will show that if x_1 and x_2 are normal words, and if $\Omega(x_1x_2) = (x_1, x_2)$, then $x_1 \geqq x_2$. This will be done by induction on the length of x_1x_2. Actually we shall prove for $n \geqq 2$ the following statement: for every linearly ordered set A, and for every pair x_1, x_2 of normal words over A with $\lambda(x_1x_2) = n$, $\Omega(x_1, x_2) = (x_1, x_2)$, we have $x_1 \geqq x_2$. If $n = 2$ this statement is true. Next we assume $n > 2$.

Let a_1, a_2 be the initial elements of x_1, x_2 respectively. Then because x_i is normal, every element of x_i is not less than a_i for $i = 1, 2$. Also, $a_1 \geqq a_2$, for if $a_1 < a_2$ we will show that x_1a_2 is normal, contradicting the assumption that $\sigma(x_1x_2) = x_1$. To show that x_1a_2 is normal if $a_1 < a_2$, we have to show that x_1a_2 is exceeded by all its shifts. First, $x_1a_2 < a_2x_1$ because $a_1 < a_2$. If $\lambda(x_1) = 1$ this shows that x_1a_2 is normal. Next, suppose $x_1 = uv$ with $u \neq \Lambda$, $v \neq \Lambda$. Then $uv < vu$ because x_1 is normal. Hence, $uva_2 < vua_2$. But the initial element of u is a_1, so $ua_2 < a_2u$. Hence, $uva_2 < vua_2 < va_2u$. This shows x_1a_2 is normal, a contradiction, so $a_1 \geqq a_2$.

If $a_1 > a_2$, we have $x_1 > x_2$ and we are done.

Finally we get to the hardest case: $a_1 = a_2 = a$. Then a is the smallest element in x_1x_2. Thus, we can put $x_1 = au_1 \cdots au_h$, $x_2 = av_1 \cdots av_j$ where the u's and v's are words either empty or with every element greater than a.

Let A_1 be the set of all $p \in A$ with $p > a$. Then words with every element greater than a are elements of A_1^*, and so is the empty word. The combinations ax with $x \in A_1^*$, will be called *syllables*. The $au_1, \cdots, au_h, av_1, \cdots, av_j$ mentioned in the previous paragraph are syllables.

Let us use the letter B for the set of all syllables. If $b_1, b_2 \in B$ we write $b_1 < b_2$ if and only if this inequality holds in A^* (elements of B are words over A). By lexicographic order, this inequality is extended to elements of B^* (the set of words over B).

There is a natural injection from words over B to words over A. For example, if $au_1, \cdots, au_k$ are syllables, then $(au_1)(au_2) \cdots (au_k)$ is a word over B, and it is mapped onto $au_1au_2 \cdots au_k$, which is a word over A. This injection is easily seen to preserve lexicographic order: $(au_1) \cdots (au_k) < (av_1) \cdots (av_j)$ in B^* if and only if $au_1 \cdots au_k < av_1 \cdots av_j$ in A^*. And it preserves normality: $(au_1) \cdots (au_k)$ is normal in B^* if and only if $au_1 \cdots au_k$ is normal in A^*.

Now let $x_1 = au_1 \cdots au_k$ and $x_2 = av_1 \cdots av_j$ be normal in A^*, with $\lambda(x_1x_2) = n$ and $\Omega(x_1x_2) = (x_1, x_2)$. Then in B^* the words $y_1 = (au_1) \cdots (au_k)$ and $y_2 = (au_1) \cdots (av_j)$ are normal, with $\Omega(y_1y_2) = (y_1, y_2)$. But $\lambda(y_1y_2)$ is less than n (the case that all syllables in x_1x_2 have length 1 is easily dismissed because $n > 2$). So by the induction hypothesis we have $y_1 \geqq y_2$. Since the injection preserves order, we conclude $x_1 \geqq x_2$. This completes the proof. □

The foregoing lemmas lead to the following.

THEOREM 1. *Every $x \in A^n$ corresponds to exactly one n-multiset f over $\mathcal{A}(A)$ such that $\Omega(x) = \nu(f)$. Furthermore $w(x) = W(f)$.*

This concludes our combinatorial proof of (5). To illustrate the bijection, consider all words over $A = \{a, b, c\}$ with $a < b < c$ which have two as, one b and one c. These words together with their Ω-factorizations are:

x	$\Omega(x)$	x	$\Omega(x)$
aabc	(*aabc*)	*baac*	(*b*, *aac*)
aacb	(*aacb*)	*baca*	(*b*, *ac*, *a*)
abac	(*abac*)	*bcaa*	(*bc*, *a*, *a*)
abca	(*abc*, *a*)	*caab*	(*c*, *aab*)
acab	(*ac*, *ab*)	*caba*	(*c*, *ab*, *a*)
acba	(*acb*, *a*)	*abaa*	(*c*, *b*, *a*, *a*)

We have listed the words in this illustration in lexicographical order. It might be noticed that if each $\Omega(x)$ is viewed as a word over $N(A)$, then the list of Ω-factorizations is also in lexicographical order. An explanation for this is given by the following results.

LEMMA 5. *Let $x, y \in A^*$ with $x \leqq y$. Then $\sigma(x) \leqq \sigma(y)$.*

Proof. If $x \leqq y$, there are two cases to consider. In the first case, we have $y = xt$ for some $t \in A^*$. Then $\sigma(x) \subseteq \sigma(xt) = \sigma(y)$, so $\sigma(x) \leqq \sigma(y)$. In the second case we have $x = upr$, $y = uqs$, $u, r, s \in A^*$, $p, q \in A$, and $p < q$. If $u = \Lambda$, then $p \subseteq \sigma(x)$, $q \subseteq \sigma(y)$, so $\sigma(x) < \sigma(y)$. If $u \neq \Lambda$, then either $\sigma(x) \subseteq u$, or $up \subseteq \sigma(x)$. If $\sigma(x) \subseteq u$, then $\sigma(x) = \sigma(u) \subseteq \sigma(y)$ since $u \subseteq y$, so $\sigma(x) \leqq \sigma(y)$. If $up \subseteq \sigma(x)$, we will show uq is normal, so $uq \subseteq \sigma(y)$, and we have $\sigma(x) < \sigma(y)$. To see that $up \subseteq \sigma(x)$ implies uq normal, suppose $\sigma(x) = upt$, $t \in A^*$. Let $u = u_1u_2$ with $u_1 \neq \Lambda$, then $u_1u_2pt < u_2ptu_1$ because upt is normal. We finally show that replacing pt by q preserves this inequality. We write $u_1u_2 = vrw$ where $v, w \in A^*$, $r \in A$ and $\lambda(v) = \lambda(u_2)$. Since $u_1u_2pt < u_2ptu$, we have $vrwpt < u_2ptu_1$. We consider the two cases $v < u_2$ and $v = u_2$ separately. If $v < u_2$ then $vrwq < u_2qu_1$ (because of $\lambda(v) = \lambda(u_2)$), so $u_1u_2q < u_2qu_1$. If $v = u_2$ we have $r \leqq p$, and therefore $r < q$, whence $vrwq < u_2qu_1$, so again $u_1u_2q < u_2qu_1$. This means that uq is normal, and our proof is complete. □

THEOREM 2. *Suppose $x < y$. Then $\Omega(x) < \Omega(y)$.*

Proof. This is proved by induction on $\lambda(x)$. The case $\lambda(x) = 1$ is trivial. Suppose the theorem is true for all $x \in A^*$ with $\lambda(x) < n$ for some $n \geqq 2$. Let $x, y \in A^*$ with $x < y$, $\lambda(x) = n$. We know from the previous lemma that $\sigma(x) \leqq \sigma(y)$. If $\sigma(x) = \sigma(y)$, then $\tau(x) < \tau(y)$, and $\lambda(\tau(x)) < \lambda(x)$, so $\Omega(\tau(x)) < \Omega(\tau(y))$ by the induction hypothesis. Hence, $\Omega(x) = (\sigma(x), \Omega(\tau(x))) < (\sigma(x), \Omega(\tau(y))) = \Omega(y)$. If $\sigma(x) < \sigma(y)$, then $\Omega(x) = (\sigma(x), \Omega(\tau(x))) < (\sigma(y), \Omega(\tau(y))) = \Omega(y)$. This completes the proof. □

As an application of Theorem 2, we mention that if v and w are normal words over A, and if $v < w$, then $v^n < w$ for all $n \in \mathbb{P}$. For if v and w are viewed as single letters, and $v < w$, then $v \cdots v$ is lexicographically less than w.

We state without proof another curious property. Let $\varepsilon(x)$ be the longest normal terminal word of x for all $x \in A^*$, and let $\delta(x)$ be the rest of x after $\varepsilon(x)$ has been deleted. Define $\Omega'(x)$, a factorization of x into normal words, as follows. First, $\Omega'(x) = (x)$ if x is normal. Otherwise if x is not empty, $\Omega'(x) = (\Omega'(\delta(x)), \varepsilon(x))$, and extra parentheses are deleted as in the definition of Ω. The surprise is that $\Omega(x) = \Omega'(x)$!

5. Algebraic proof of (5). Recall that $P_1(A)$ is the set of aperiodic words over A, and that $\mathscr{A}(A)$ is the set of aperiodic cycles over A. Every aperiodic cycle of length n can give rise to n distinct aperiodic words (which are obtained by breaking the cycle open at one of the n possible places).

Because of the definition of W and w in § 2 we have

$$\sum_{f \in \mathscr{M}_1(A)} W(f) = \sum_{C \in \mathscr{A}(A)} \prod_{k=0}^{\infty} (w(C))^k. \tag{10}$$

We shall use the identity

$$\sum_{k=0}^{\infty} z^k = \exp\left\{\sum_{k=1}^{\infty} \frac{z^k}{k}\right\}. \tag{11}$$

Applying this with $z = w(C)$, (10) becomes

$$\sum_{f \in \mathscr{M}_1(A)} W(f) = \prod_{C \in \mathscr{A}(A)} \exp\left\{\sum_{k=1}^{\infty} \frac{(w(C))^k}{k}\right\} = \exp\left\{\sum_{k=1}^{\infty} \frac{1}{k} \sum_{C \in \mathscr{A}(A)} (w(C))^k\right\}. \tag{12}$$

We shall prove the identity

$$\sum_{k=1}^{\infty} \frac{1}{k} \sum_{C \in \mathscr{A}(A)} (w(C))^k = \sum_{m=1}^{\infty} \frac{1}{m} \sum_{y \in A^m} w(y). \tag{13}$$

The left-hand side can be written as

$$\sum_{k=1}^{\infty} \sum_{n=1}^{\infty} \frac{1}{kn} \sum_{\substack{x \in P_1(A) \\ \lambda(x)=n}} (w(x))^k, \tag{14}$$

since each aperiodic cycle C with length n corresponds to exactly n elements of $P_1(A)$.

Taking terms with the same value of kn together, we transform (14) into

$$\sum_{m=1}^{\infty} \frac{1}{m} \sum_{n|m} \sum_{\substack{x \in P_1(A) \\ \lambda(x)=n}} (w(x))^{m/n}. \tag{15}$$

Every word y with length m can be written uniquely as $y = x^{m/n}$, where n is a divisor of m and x is aperiodic. Conversely, if n divides m and if $x \in P_1(A)$, $\lambda(x) = n$, then $x^{m/n} \in A^m$. Therefore we can write (15) as the right-hand side of (13), just noting that $(w(x))^{m/n} = w(x^{m/n})$. This proves (13).

Since

$$\sum_{y \in A^n} w(y) = \left(\sum_{a \in A} w(a) \right)^n \tag{16}$$

we find that application of the exponential function to the right-hand side of (13) leads to

$$\sum_{k=0}^{\infty} \left(\sum_{a \in A} w(a) \right)^k$$

(cf. (11)), whence (12) and (13) lead to (5).

As a generalization of (13) we mention, with an extra parameter s,

$$\sum_{k=1}^{\infty} \frac{1}{k^{s+1}} \sum_{c \in \mathscr{A}(A)} \frac{(w(c))^k}{(\lambda(c))^s} = \sum_{y \in A^* \setminus \{\Lambda\}} \frac{w(y)}{(\lambda(y))^{s+1}}. \tag{17}$$

The case $s = 0$ is (13), but the case $s = -1$ looks pretty as well.

6. Some examples. Let us return to the problem we described in § 3. Let D be an n-set, $n \in \mathbb{P}$, let $S(D)$ be the set (and group) of all permutations of D, and let D^D be the set of all mappings of D into D. Elements of D^D are called *endomorphisms* of D. The (directed) *graph* of $f \in D^D$ has vertex set D and edge set $\{(d, f(d)): d \in D\}$. Elements $f, g \in D^D$ are defined to be *equivalent* if the graph of f is isomorphic to the graph of g. This means there exists $\gamma \in S(D)$ such that $\{(\gamma d, \gamma f(d)): d \in D\} = \{(d, g(d): d \in D\} = \{(\gamma d, g\gamma(d): d \in D\}$, that is, $\gamma f = g\gamma$, which is the same as $\gamma f \gamma^{-1} = g$. An equivalence class in D^D is called an *endomorphism pattern* of D, and $f(n)$ is defined to be the number of these patterns for any n-set D. Next, $t(n)$ is defined to be the number of isomorphic classes of rooted trees with vertex set any n-set V, $n \in \mathbb{P}$. Such a class is called a *rooted tree pattern*. Diagrams representing rooted tree patterns having fewer than six vertices are shown in Fig. 1. Diagrams representing endomorphism patterns of an n-set for $n = 1, 2, 3$ are shown in Fig. 2. Also, Fig. 3 indicates how an endomorphism pattern might be viewed as a multiset of cycles of rooted tree patterns.

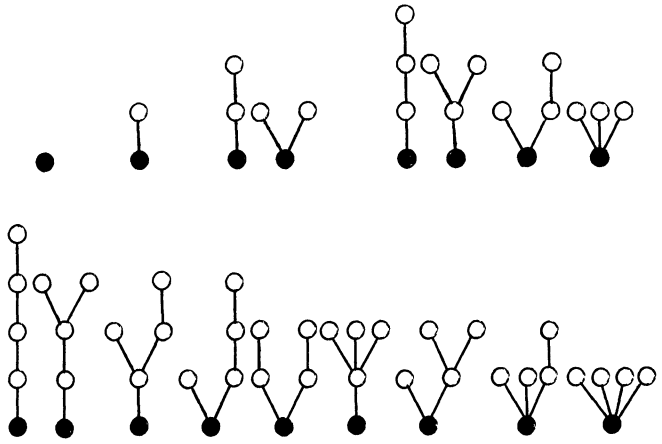

FIG. 1. *Rooted tree patterns.*

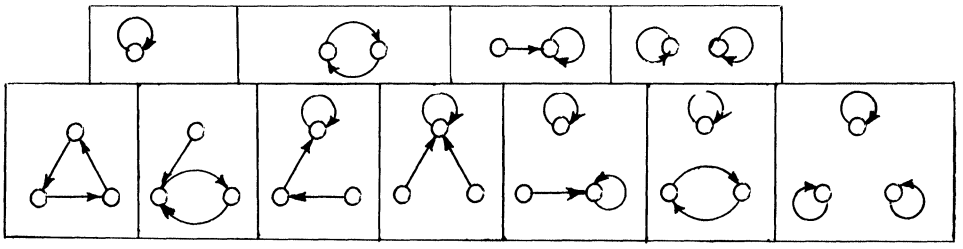

FIG. 2. *Endomorphism patterns.*

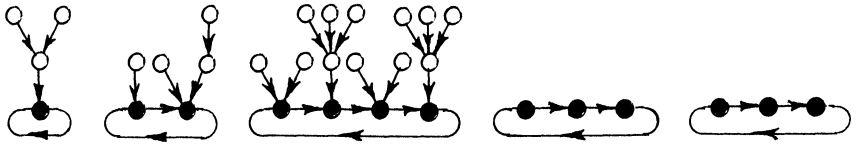

FIG. 3. *An endomorphism pattern viewed as a multiset of rooted tree cycles.*

Read's formula (8) tells us how to compute $f(n)$, the number of endomorphism patterns on an n-set, in terms of the rooted tree numbers. However, the bijection described in § 4 tells how to encode such a pattern as a sequence of words over the set of trees. The kth word in the sequence corresponds to the k-fold cycles of trees. For example, the endomorphism pattern in Fig. 3 is encoded as (6, 2, 7), (4, 14), (1, 1), $(\Lambda, \Lambda, \Lambda, \cdots)$. To decode a sequence of words $(x_1, x_2, \cdots)$, apply Ω to x_k for $k = 1, 2, \cdots$ to get $\Omega(x_k) = (x_{k1}, x_{k2}, \cdots)$; then x_{ki} is used to form a k-fold cycle of rooted trees $\langle x_{ki}^k \rangle$. We use as a convention that x_k is replaced by Λ if there are no k-fold cycles.

Our generalization of Read's formula allows us to enumerate other kinds of endomorphism patterns with equal ease. For example, suppose we are only interested in those $f \in D^D$ with $|f^{-1}(d)| \leqq h$; that is, every $d \in D$ is the image under f of at most h other elements of D, $h \in \mathbb{P}$. When $h = 1$, these endomorphisms are the permutations of D, and the patterns correspond to partitions of n if D is an n-set. For any $h \in \mathbb{P}$, endomorphisms f such that $|f^{-1}(d)| \leqq h$ give rise to patterns whose encoding as trees involves a special sort of rooted tree. Namely, the in-degree of the root vertex is at

most $h-1$, and the in-degrees of all other vertices are at most h. When $h=1$, there is only one tree like this, namely, the rooted tree with one vertex. In this case we would take $T(z)=z$ in (8) to get

$$\sum_{n=0}^{\infty} f(n)z^n = \prod_{k=1}^{\infty} \frac{1}{1-z^k}, \tag{18}$$

which is the generating function for the number of partitions of n, as expected. When $h=2$, the form of T is in part

$$T(z)=z+z^2+z^3+2z^4+3z^5+6z^6+11z^7+23z^8+46z^9+98z^{10}+\cdots.$$

We note that $T(z)=z(1+S(z))$, where $S(z)$ is the generating function for the rooted trees in which all vertices have degree $\leqq 2$. By Pólya's method (see [7]) we have

$$S(z)=z(1+S(z)+\tfrac{1}{2}S(z^2)+\tfrac{1}{2}(S(z))^2),$$

and therefore

$$T(z)=z+\tfrac{1}{2}(T(z^2)+(T(z))^2).$$

Using (8) with this new generating function T we get

$$\begin{aligned} F(z)=1&+z+3z^2+6z^3+15z^4+31z^5+75z^6+164z^7\\ &+388z^8+887z^9+2092z^{10}+4884z^{11}+11599z^{12}+27443z^{13}\\ &+65509z^{14}+156427z^{15}+375263z^{16}+\cdots. \end{aligned}$$

Thus, there are exactly 887 endomorphism patterns on a 9-set, involving $f\in D^D$ such that $|f^{-1}(d)|\leqq 2$ for all $d\in D$.

We close with some comments on the several papers which have dealt with the computation of $f(n)$. Fisher (1942) [4] seems to be first, and the same article with some corrections and additions appears in [5] (1950). In his 1950 reprinting of his earlier paper, Fisher adds a note indicating that he was unaware of Pólya's enumeration method. Nevertheless, Fisher's method produces results which run parallel to what one would get using Pólya's method. Davis (1953) [3] was aware of Pólya's method, but he elected to give an explicit formula for $f(n)$ using "Burnside's lemma" which is now properly renamed the Cauchy–Frobenius theorem. (See de Bruijn [1].) Harary (1959) [6] touched on the problem of computing $F(z)$ in his enumeration of patterns of functional digraphs, and he used Pólya's method. Read (1959) [8] obtained (8) by simplifying a formula given in Harary's paper. Finally, de Bruijn (1972) [2] investigated endomorphism patterns using the group action $\rho(\gamma)f=\gamma f\gamma^{-1}$, finding new proofs for older results.

Note added in proof. We are indebted to D. Foata for pointing out that Theorem 1 was known in the context of the theory of free Lie algebras. The oldest reference seems to be to A. I. Širšov, *Subalgebras of free Lie algebras*, Mat. Sbornik N.S. 33 (75) (1953) pp. 441–452. For related results see G. Viennot, *Algèbres de Lie libres et monoïdes libres*, Lecture Notes in Mathematics 691, Springer-Verlag, Berlin, Heidelberg, New York, 1978.

REFERENCES

[1] N. G. DE BRUIJN, *A note on the Cauchy–Frobenius lemma*, Nederl. Akad. Wetensch. Proc. Ser. A, 82 (= Indag. Math., 41) (1979), pp. 225–228.

[2] ———, *Enumeration of mapping patterns*, J. Combin. Theory, 12 (1972), pp. 14–20.
[3] R. L. DAVIS, *The number of structures of finite relations*, Proc. Amer. Math. Soc., 4 (1953), pp. 486–495.
[4] R. A. FISHER, *Some combinatorial theorems and enumerations connected with the number of diagonal types of a latin square*, Ann. Eugenics, XI, pt. IV, (1942), pp. 395–401.
[5] ———, *Contributions to Mathematical Statistics*, John Wiley, New York, 1950, pp. 41.394a–41.401.
[6] F. HARARY, *The number of functional digraphs*, Math. Ann., 138 (1959), pp. 203–210.
[7] G. PÓLYA, *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen*, Acta Math., 68 (1937), pp. 145–254.
[8] R. C. READ, *A note on the number of functional digraphs*, Math. Ann., 143 (1961), pp. 109–110.