PACIFIC JOURNAL OF MATHEMATICS
Vol. 85, No. 2, 1979

AVOIDABLE PATTERNS IN STRINGS OF SYMBOLS

DWIGHT R. BEAN, ANDRZEJ EHRENFEUCHT AND GEORGE F. MCNULTY

A *word* is just a finite string of letters. The word W *avoids* the word U provided no substitution instance of U is a subword of W. W is *avoidable* if on some finite alphabet there is an infinite collection of words each of which avoids W. W is kth *power-free* if W avoids x , where x is a letter. We develope the theory of those endomorphisms of free semigroups which preserve kth power-freeness and employ this theory to investigate kth power-free words. We go on to prove that every kth power-free word on n letters is a subword of a maximal word of the same kind. Next we examine avoidable words in general and prove that all words of length at least 2^n on an alphabet with n letters are simultaneously avoidable. We show that on any finite alphabet the collection of avoidable words is simultaneously avoidable. We provide an effective (recursive) characterization of avoidability. Finally we show how our work can be extended to infinite words, to n-dimensional arrays, and to circular words. We give an application to the Burnside problem for semigroups. The present work is chiefly concerned with certain combinatorial properties of strings of symbols. As such, it belongs to formal linguistics, to the theory of free semigroups, and to the theory of partitioned linear orders. While we have taken all of these points of view in the body of this work, it has proven most convenient to base our exposition on an attitude between linguistics and free semigroups.

By an *alphabet* we mean any set, the members of which are called *letters* and can be regarded in all subsequent discussions as indivisible. A *word on* the alphabet N is a finite string of letters belonging to N. For example, if $N = \{a, b, c, d\}$ then $abacd$ is a word on N. The *empty word* is the string with no letters and it is regarded as a word on every alphabet. Words can be concatenated: whenever U and V are words, the result of concatenating U and V is expressed by juxtaposition. If $U = abacd$ and $V = bdaca$, then $UV = abacdbdaca$. If $W = UV$ then U is an *initial segment* of W and V is a *final segment*. U is a *subword* of W provided $W = XUY$ for some words X and Y. For any word W and any natural number k, W^k is defined so that

$$W^0 \text{ is the empty word}$$
$$W^{k+1} = W^k W .$$