Avoidable words and lattice universal semigroup varieties

by

Pavel Goralcik

The "purely speculative field" opened to mathematical research by Axel Thue [7,8] about 80 years ago - combinatorics of words - has found many applications since. The original problem of avoidance of squares in arbitrarily long words on a small (ternary) alphabet continues to be a source of inspiration. The notion was recently generalized by Bean, Ehrenfeucht, and McNulty [1] to the problem of avoidance of an arbitrary word. In this form the problem perfectly fits into the framework of universal algebra. The idea goes back to the application of square-free words to semigroup varieties by Burris, Nelson [3], and Jezek [5].

It is well known, since the work of G. Birkhoff, that the classes of semigroups closed under the passage to homomorphic images, subsemigroups, and direct products - the so called varieties - stand in one-one correspondence with the sets of equations satisfied by them. Recall that by an *equation* is meant any pair (u,v) of words over a countably infinite standard alphabet X; such an equation is satisfied by a class K of semigroups if $f(u)=f(v)$ for every homomorphism $f:X^+ \longrightarrow S$ with S in K. The set of all equations satisfied by K is called the *equational theory* of K and can be described as

$$Th(K) = \{Ker\ f;\ f \in Hom(X^+,S),\ S \in K\}$$

Algebraically, Th(K) is a *fully invariant congruence* in the free semigroup X^+, i.e. a congruence closed to endomorphisms:

$$(u,v) \in Th(K) \text{ implies } (f(u),f(v)) \in Th(K) \text{ for all } f \in End\ X^+.$$

The one-one correspondence between the semigroup varieties **V** on the one hand and their equational theories Th(**V**) on the other hand is known to be a dual lattice isomorphism.

The variety Var(K) generated by a class K of semigroups can be considered as an adequate expression of the "amount of algebraic structure contained in K". We can say that a class K has at most as much semigroup structure as another class K' provided Var(K) $\subseteq$ Var(K'). In particular, any class K or a single semigroup S which generates the variety Sem of all semigroups is endowed, in this sense, with a maximum possible amount of semigroup structure. In terms of equations, no non-trivial equation (u,v), u$\neq$v, is satisfied by such a K or S.

The structural hierarchy of classes of semigroups is thus expressed by the lattice L_{Sem} of semigroup varieties. Each variety **V** has its own hierarchy in the form of the lattice L_V of subvarieties of **V**.

Now, applying the same structural approach to the lattice L_V we come to a fairly sophisticated way of comparing semigroup varieties **V** by the amount of lattice structure contained in L_V. Naturally distinguished then become those varieties **V** whose lattice L_V does not satisfy any non-trivial lattice equations. Let us call such varieties **V** *lattice universal*.

How to prove that a given lattice L does not satisfy any non-trivial equation? One feasible way is to embed into L the partition lattice Π_∞ over a countably infinite set since this lattice has been proved by Sachs [6] not to satisfy any non-trivial lattice equations.

The lattice L_V of subvarieties of a semigroup variety **V** is dually isomorphic to the interval of equational theories between Th(**V**) and the total congruence Th($\mathbb{T}$), the equational theory of the variety **T** of trivial semigroups. Therefore we have

Statement 1. If a semigroup variety **V** is such that the partition lattice on a countably infinite set Π_∞ can be embedded into the interval [Th(**V**),Th(**T**)] of equational theories, then **V** is lattice universal.

All told, let us now look more closely at congruences in X^+. Among the simplest ones are the so called *Rees congruences*, defined by ideals I in X^+ as the congruence generated by $I \times I$. The corresponding partition of X^+ has I for the only non-singleton class. When a Rees congruence is a theory? Obviously, if and only if the ideal I defining it is invariant under the endomorphisms of X^+, $f(I) \subseteq I$ for all $f \in \mathrm{End}\ X^+$.

The invariant ideals in X^+ form a closure system $\mathcal{Y}$ easily described as follows. For a word w in X^+, denote by [w] the set of all endomorphic images ("substitution instances") of w,

$$[w] = \{f(w);\ f \in \mathrm{End}\ X^+\}$$

Then we have

$$\mathcal{Y}(w) = X^*[w]X^*, \qquad \mathcal{Y}(W) = \bigcup\{\mathcal{Y}(w);\ w \in W\}$$

for any $W \subseteq X^+$.

Combinatorially speaking, $\mathcal{J}(w)$ consists of the words on X containing as a factor a substitution instance of the word w. The complement of $\mathcal{J}(w)$ in X^+ consists of the words u having no substitution instance of w as a factor; such a word u is said to *avoid* the word w. We have come to the basic concept in [1].

Definition 1. A word w in X^+ is *avoidable on a finite alphabet* $A \subseteq X$ if there is an infinite collection of words on A avoiding w; the word w is *avoidable* if it is avoidable on some finite alphabet A.

This definition is a direct generalization of the avoidance of squares: the word aa is avoidable on the ternary alphabet $\{a,b,c\}$. Our previous considerations lead us to considerably strengthen the notion of avoidance.

Definition 2. An infinite set of words J X^+ is a *Jezek set* if

(1) all w in J have the same set A of letters,

(2) J is $\mathcal{J}$-independent, in the sense that $u \notin \mathcal{J}(v)$ for any pair of distinct $u,v \in J$.

A word w X^+ is *strongly avoidable* if there is a Jezek set $J \subseteq X^+$ of words avoiding w.

Statement 2. If the word $w \in X^+$ is strongly avoidable then the variety W determined by the Rees theory associated with the ideal $\mathcal{J}(w)$ is lattice universal. The theory Th(W) is generated by the equations $wx=xw=w$ (saying that $w/Th(W)$ is a zero in $X^+/Th(W)$).

A sketch of proof runs as follows. Let J be a Jezek set avoiding w, let A be the set of letters occurring in the words of J. Let K denote the fully invariant ideal

$$X^*([w] \cup [J])X^* \setminus \{f(u);\ u \in J,\ f \in \mathrm{Aut}\ X^+\}$$

Assign to each equivalence E on J the equational theory generated by $E \cup (K \times K)$. This assignment is an embedding of the lattice of equivalences on J (isomorphic to Π_∞) onto an interval of the lattice of equational theories.

Jezek [5] is the first one to have constructed a Jezek set of square-free words on the ternary alphabet {a,b,c}. His construction enabled Goralcik and Vanicek [4] to produce, via encodings of the original Jezek set J, to every binary word w on A = {a,b} avoidable on A a Jezek set J_w avoiding w.

Theorem. Let A = {a,b}. A word $w \in A^+$ is avoidable on A if and only if $w \in A^*[T]A^*$, where

$$T = \{aaa,\ ababa,\ abaab,\ abbab,\ aabba,\ abbaa,\ aababb\}.$$

The complement $A^+ \setminus A^*[T]A^*$ consists of 28 words each of which is unavoidable on A.

Every word $w \in A^+$ which is avoidable on A is strongly avoidable on A.

To prove the first assertion, we only need to construct a Jezek set J_w for w from the set of five words

aaa, ababa, aabba, aababb, abaab

(because the remaining two in T are isomorphic to the reversals of aabba and abaab). The required Jezek sets are obtained by encoding a fixed Jezek set J of square-free words on {a,b,c} by the following codes:

For aaa:	abbababbabaabaabb
	abbabaababbabaabb
	abbaababbababbaabb
For ababa:	ababbaabbb
	abbaababbb
	aababbabbb
For aabba:	abbaaabaabababaaaabaabbbabbbb
	abbaaabaaaabaaababbbb
	abbaaabaaabaababababaaaabbbabbbb
For aababb:	abbaaaaabaaaaabaaaaabaaaaabbb
	abbaaaabaaaabaaaabaaaabbb
	abbaaabaaabaaabaaabaaabbb
For abaab:	aaaababababaaaababbbb
	ababababaaaabababbbbb
	abaaaabababababaaabbbb

The verification that it works is tedious.

One curious consequence of the above encoding is that for each w in T the rate of growth of the number of words avoiding w on A={a,b} (as a function of length) is the same as the growth of the number of square-free ternary words, which is exponential by Brandenburg [2].

References

[1] D. Bean, A. Ehrenfeucht, and G. McNulty, Avoidable patterns in strings of symbols, *Pacific J. Math.* **84**(1979)261-294.

[2] F. Brandenburg, Uniformly growing k-th power-free homomorphisms, *Theor. Comput. Sci.* **23**(1983)69-82.

[3] S, Burris, E. Nelson, Embedding the dual of in the lattice of equational classes of semigroups, *Algebra Universalis* **2**(1971)248-253.

[4] P. Goralcik, T. Vanicek, Binary patterns in binary words, to appear.

[5] J. Jezek, Intervals in the lattice of varieties, *Algebra Universalis* **6**(1976)147-158.

[6] D. Sachs, Identities in finite partition lattices, *Proc. Amer. Math. Soc.* **12**(1969)944-945.

[7] A. Thue, Uber unendliche Zeichenreihen, *Norske Vid. Selsk. Skr.* I, *Math. Nat. Kl. Christiania* VII(1906)1-22.

[8] A. Thue, Uber die gegenseitigen Lage gleicher Teile gewisser Zeichenreihen, *Norske Vid. Selsk. Skr.* I *Math. Nat. Kl. Christiania* I(1912)1-67.

[9] T. Vanicek, Unavoidable words (in Czech), Diploma Thesis, Faculty of Mathematics and Physics, Charles University, 1989.

Pavel Goralcik
MFF KU, Sokolovska 83
186 00 Praha 8 (Czechoslovakia)