HYPERCODES, RIGHT CONVEX LANGUAGES
AND THEIR SYNTACTIC MONOIDS

G. THIERRIN

Abstract. If $X^*$ is the free monoid generated by the alphabet $X$, then any subset $L$ of $X^*$ is called a language over $X$. If $P_L$ is the principal congruence determined by $L$, then the quotient monoid $\text{syn}(L) = X^*/P_L$ is called the syntactic monoid of $L$. A hypercode over $X$ is any set of nonempty words that are noncomparable with respect to the embedding order of $X^*$. If $H$ is a hypercode, then the language $H' = \{x | x \in X^* \text{ and } a < x \text{ for some } a \in H\}$ is a right convex ideal of $X^*$. The syntactic monoid $\text{syn}(H)$ can be characterized as a monoid with a disjunctive $\mu$-zero. The two particular interesting cases when $\text{syn}(H)$ is a nil monoid and when $\text{syn}(H)$ is a semilattice are also characterized.

1. Introduction and preliminary results. Let $X$ be an alphabet, finite or infinite, let $X^*$ be the free monoid generated by $X$ and let $X^+ = X^* - \{1\}$, $1$ being the empty word. The length of a word, $x \in X^*$, is denoted by $\lg(x)$ and every subset of $X^*$ is called a language over $X$.

If $M$ is a monoid and $A$ is a subset of $M$, then the relation $P_A$, defined by $a \equiv b(P_A)$ iff $A.a = A.b$ where $A.a = \{(x, y) | x, y \in M, xay \in A\}$ is a congruence of $M$, called the principal congruence determined by $A$. If $P_A$ is the identity, then $A$ is called disjunctive.

If $A$ is a language over $X$, then $P_A$ is called the syntactic congruence of $A$ and the quotient monoid $\text{syn}(A) = X^*/P_A$ is called the syntactic monoid of $A$.

The relation $\prec$ defined on $X^*$ by $x \prec y$ iff $x = x_1x_2 \cdots x_n$ and $y = y_1y_2y_3 \cdots y_ny_{n+1}$ for some $n$ where $x_i, y_i \in X^*$ is a partial order on $X^*$, called the embedding order. For every language $A \subseteq X^*$, let $\tilde{A} = \{x | a < x \text{ for some } a \in A\}$ and $A = \{x | x < a \text{ for some } a \in A\}$. It is well known that if $X$ is finite, then each set of pairwise incomparable elements of $X^*$ is always finite ([1], [2], [3]), and that the languages $\tilde{A}$ and $A$ are regular, that is, their syntactic monoids are finite.

A nonempty language, $C \subseteq X^+$, is called a code if $a_1a_2 \cdots a_n = b_1b_2 \cdots b_m$ and $a_i, b_j \in C$ implies $n = m$ and $a_i = b_i$ for every $i$. A nonempty language $H \subseteq X^+$ is called a hypercode if every pair of distinct elements of $H$ are incomparable relative to the embedding order. It is immediate that every hypercode is a code.

Received by the editors June 28, 1980 and, in revised form, January 13, 1981.
1980 Mathematics Subject Classification. Primary 20M35; Secondary 68A30.

Key words and phrases. Monoid, language, syntactic monoid, embedding order, hypercode, convex, nil monoid, semilattice.

This research has been supported by Grant A7877 of the Natural Sciences and Engineering Research Council of Canada.
A language $A \subseteq X^*$ is said to be right convex if $a < x$, $a \in A$ implies $x \in A$. A language $A$, $1 \notin A$, is right convex iff $A = \bar{H}$ for some hypercode $H$. This hypercode $H$ is the set of the minimal words of $A$. Right convex languages and hypercodes have been first considered when the alphabet $X$ is finite ([7], [8]). In this case, the hypercodes are always finite and the right convex languages are regular.

An ideal $I$ of a monoid $M$ is called a $\mu$-ideal if $ab \in I$ implies $axb \in I$ for all $x \in M$ [10] and a zero element of $M$ is called a $\mu$-zero if it is a $\mu$-ideal. A language $A$ over $X$ is right convex iff $A$ is a $\mu$-ideal.

If $H$ is a hypercode over $X$, then $\text{syn}(\bar{H})$ is a monoid with a disjunctive $\mu$-zero. Conversely, if $M$ is a monoid with a disjunctive $\mu$-zero, then there exists a hypercode over an alphabet $X$ such that $M$ is isomorphic to $\text{syn}(\bar{H})$.

Some properties of a language can be determined by considering their syntactic monoid. For example, a language $A$ over a finite alphabet is regular iff $\text{syn}(A)$ is finite. If the language $C$ is a code, then it is in general more interesting to consider the syntactic monoid $\text{syn}(C^*)$ of $C^*$ instead of $C$. However, this is not the case for the hypercodes. A characterization of the syntactic monoid $\text{syn}(H)$ of a hypercode $H$ over a finite alphabet has been given in [9]. In this paper, we consider the syntactic monoid of languages of the form $\bar{H}$ where $H$ is a hypercode. In particular, we characterize the hypercodes $H$ in the two following cases:

(a) $\text{syn}(\bar{H})$ is a nil monoid and the alphabet $X$ is finite;
(b) $\text{syn}(\bar{H})$ is a semilattice.

2. Quasi-maximal hypercodes over a finite alphabet. In this section, the alphabet $X$ is always assumed to be finite. A hypercode $H$ over $X$ is said to be maximal if for every, $u \in X^*$, $u \notin H$, $H \cup u$ is not a hypercode. Every hypercode can be embedded in a maximal one. A hypercode is maximal iff $X^* = \bar{H} \cup H$ [7]. A hypercode is said to be quasi-maximal if $X^* - (\bar{H} \cup H)$ is finite. Since the alphabet $X$ is finite, and since a hypercode over $X$ is always finite, then $\bar{H}$ is also finite; therefore, $H$ is a quasi-maximal hypercode iff $X^* - \bar{H}$ is finite. Clearly, every maximal hypercode is quasi-maximal, but the converse is not true. For example, if $X = \{a, b\}$, then $H = \{a^2, b^2\}$ is a hypercode that is quasi-maximal but not maximal.

Let us remark that a hypercode $H$ over a finite alphabet $X$ is quasi-maximal iff there exists an integer $m > 1$ such that $H \cup u$ is not a hypercode for $\log(u) > m$.

Recall that a nil monoid is a monoid with zero such that every element different from the identity is nilpotent.

If $A$ is a language over $X$ we will denote by $\alpha(A)$ the alphabet of $A$, that is, $\alpha(A) = \{a | a \in X \text{ and } ras \in A \text{ for some } r, s \in X^*\}$.

**Proposition 1.** Let $H$ be a hypercode over a finite alphabet $X$ such that $\alpha(H) = X$. Then $H$ is quasi-maximal $\iff$ $\text{syn}(\bar{H})$ is a finite nil monoid.

**Proof.** ($\Rightarrow$) Let $u \in X^*$, $u \neq 1$, and suppose that $u^m \notin \bar{H}$ for $m > 1$. Since $H$ is quasi-maximal, there exists $k > 1$ such that $H \cup u^n$ is not a hypercode for $n > k$. Hence $u^n \in \bar{H}$ for $n > k$. Since $\bar{H}$ is finite, we have a contradiction. Therefore,
um \in \tilde{H} \text{ for some } m > 1 \text{ and syn}(\tilde{H}) \text{ is a nil monoid because the class } \tilde{H} \text{ modulo } P_{\tilde{H}} \text{ is the zero element of syn}(\tilde{H}). \text{ Since } H \text{ is finite, then } \tilde{H} \text{ is regular and syn}(\tilde{H}) \text{ is finite.}

(\Leftarrow) \text{ Let } a \in X. \text{ Since } a(H) = X, \text{ there exist } r, s \in X^* \text{ such that } h_1 = ras \in H. \text{ If } a \equiv 1(P_{\tilde{H}}), \text{ then } h_1 = ras \equiv rs(P_{\tilde{H}}). \text{ Since } \tilde{H} \text{ is a class modulo } P_{\tilde{H}}, \text{ then } rs \in \tilde{H} \text{ and } h_2 < rs \text{ for some } h_2 \in H. \text{ Therefore, } h_2 < rs < h_1 \text{ and } h_2 = rs = h_1, \text{ a contradiction. Hence } a \not\equiv 1(P_{\tilde{H}}).

Since syn(\tilde{H}) \text{ is a finite nil monoid, then for every } a \in X, \text{ there exists } m > 1 \text{ such that } a^m \in \tilde{H}. \text{ Suppose that } X^* - (\tilde{H} \cup \tilde{H}) = K \text{ is infinite. Then there exists } w \in K \text{ containing at least } m \text{ identical letters of the alphabet } X, \text{ say } a. \text{ Therefore, } w = x_1ax_2a \cdots ax_max_{m+1} \text{ where } x_i \in X^*. \text{ Since } a^m \in \tilde{H}, \text{ then } a^m < w \text{ and } w \in \tilde{H}, \text{ a contradiction. It follows then that } H \text{ is quasi-maximal.} \quad \square

Remark that if } H \text{ is a quasi-maximal hypercode over a finite alphabet, then syn}(\tilde{H}) \text{ is a nil monoid with a disjunctive } \mu\text{-zero and syn}(\tilde{H}) \text{ is subdirectly irreducible, because every nil monoid with a disjunctive zero is subdirectly irreducible [6].}

3. Hypercodes and semilattices with disjunctive zero. In this section, we characterize the hypercodes } H \text{ such that syn}(\tilde{H}) \text{ is a semilattice with a disjunctive zero. Let us remark that if a semilattice has a zero, then the zero element is disjunctive iff the sets of the zero divisors of two distinct elements are always distinct.}

If } B(V, \land) \text{ is a boolean algebra, then the zero element 0 of } B \text{ (the identity element 1 of } B) \text{ is a disjunctive zero relative to the operation } \land \text{ (the operation } \lor). \text{ A semilattice or a lattice with a disjunctive zero is not in general a boolean algebra. For example, let } L = \{0, 1, a_1, a_2, a_3\} \text{ with } a_i \land a_j = 0 \text{ and } a_i \lor a_j = 1 \text{ for } i \neq j. \text{ Clearly, 0 is a disjunctive element for the operation } \land, \text{ but } L \text{ is not a boolean algebra. Conditions for a subset, and, in particular, for an element of a semilattice to be disjunctive can be found in [4].}

Recall that an ideal } A \text{ of a monoid } M \text{ is said to be cs-prime (or completely semiprime) if } x^n \in A, \text{ } n \text{ a positive integer, implies } x \in A. \text{ Remark that an ideal } A \text{ of a monoid } M \text{ is cs-prime iff the quotient monoid } M/P_A \text{ is a semilattice with a disjunctive zero.}

Let } A \text{ be a language over } X. \text{ } A \text{ is said to be power free if } xa^n y \in A \text{ with } a \neq 1, \text{ } x, y \in X^*, \text{ } n > 1, \text{ implies } n = 1. \text{ } A \text{ is said to be completely reflective if } uwv \in H \text{ with } u, v, w \in X^* \text{ implies } wuv \in H.

**Proposition 2.** Let } H \text{ be a hypercode. Then the following properties are equivalent.

(1) \tilde{H} \text{ is cs-prime.}

(2) H \text{ is power free and completely reflective.}

(3) syn(\tilde{H}) \text{ is a semilattice with disjunctive zero.}

(4) syn(\tilde{H}) \text{ is a semilattice.}

**Proof.** (1) ⇒ (2). Suppose that } uw^3w \in H, \text{ } v \neq 1. \text{ Then } uw \cdot wu \cdot wv \in \tilde{H}, \text{ } (uwv)^2 \in \tilde{H} \text{ and } uvw \in \tilde{H}. \text{ Hence } h < uow \text{ for some } h \in H. \text{ From } h < uow \leq uw^3w \text{ it follows that } h = uow = uw^3w, \text{ a contradiction. Therefore, } H \text{ is power free.
Let \( m \) be the length of the words of minimal length in \( H \) and let \( uow \in H \) with \( \lg(uow) = m \). Then \( wvu \in \bar{H} \) because, by the above remark, \( \text{syn}(\bar{H}) \) is commutative. Hence, \( h < wvu \) for some \( h \in H \), and, therefore, \( h = wvu \in H \). Suppose now that for all words \( uow \) of length \( \lg(uow) < n \), \( wov \in H \) implies \( wvu \in H \). Let \( uow \in H \) with \( \lg(uow) = k > n \) such that for every \( r \in H \) with \( \lg(r) < k \) we have \( \lg(r) < n \). Since \( \text{syn}(\bar{H}) \) is commutative, then \( wvu \in \bar{H} \) and there exists \( h \in H \) such that \( h < wvu \). Suppose that \( wvu \notin H \). Then \( \lg(h) < \lg(wvu) \) and \( \lg(h) < n \). Furthermore, \( h = w_1v_1u_1 \) with \( w_1 < w, v_1 < v, u_1 < u \). Since \( \lg(h) < n \), then \( u_1v_1w_1 \in H \). But \( u_1v_1w_1 < uow \), and, therefore, \( u_1v_1w_1 = uow \), a contradiction. Hence, \( wvu \in H \) and \( H \) is completely reflective.

(2) \( \Rightarrow \) (1). Since \( \bar{H} \) is an ideal, in order to show that \( \bar{H} \) is cs-prime, it is sufficient to show that \( u^2 \in \bar{H} \) implies \( u \in H \). Suppose that \( u \notin \bar{H} \). Then there exists \( v \in H \) such that \( v < u^2 \) and \( v \not< u \). Let \( v = v_1v_2 \cdots v_m \) with \( v_i \in X \). Since \( H \) is completely reflective, then \( v_1v_2 \cdots v_m \in H \) for every permutation \( i_1i_2 \cdots i_m \) of \( 1, 2, \ldots, m \). Since \( H \) is power free, then \( v_j \neq v_i \) for \( i \neq j \). From \( v < u^2 \), we have \( v = v_1 \cdots v_rv_{r+1} \cdots v_m \) with \( v_1v_2 \cdots v_r < u \) and \( v_{r+1} \cdots v_m < u \). It follows then that \( u = x_1v_1x_2v_2 \cdots x_mv_mx_{m+1} \) with \( x_j \in X^* \) and \( v' = v_1v_2 \cdots v_m \) is obtained from \( v \) by a permutation of its letters. But \( v' \in H \). Therefore, \( v' < u \) and \( u \notin \bar{H} \), a contradiction.

(1) \( \Leftrightarrow \) (3). Immediate.

(3) \( \Rightarrow \) (4). Trivial.

(4) \( \Rightarrow \) (3). \( \bar{H} \) is a class modulo \( P_{\bar{H}} \), and since \( \bar{H} \) is an ideal, then \( \bar{H} \) is a disjunctive zero of \( \text{syn}(\bar{H}) \). □

Remark that in Proposition 2 the semilattice \( \text{syn}(\bar{H}) \) always has an identity element. It is immediate that if \( S \) is a semilattice with a disjunctive zero and an identity element, then \( S \) is isomorphic to \( \text{syn}(\bar{H}) \), where \( H \) is a power free and completely reflective hypercode \( H \) over some alphabet \( X \).

REFERENCES


DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WESTERN ONTARIO, LONDON, ONTARIO N6A 5B9, CANADA