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 $\leq$ defined on $X^*$ by $x \leq y$ iff $x = x_1x_2\cdots x_n$ and $y = y_1y_2y_3\cdots y_nx_{n+1}$ 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 $\bar{A} = \{x|a < x \text{ for some } a \in A\}$ and $\bar{A} = \{x|x < a \text{ for some } a \in A\}$. 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 $\bar{A}$ and $\bar{A}$ 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.
A language $A \subseteq X^*$ is said to be right convex if $a < x, a \in A$ implies $x \in A$. A language $A$, $1 \not\in A$, is right convex iff $A = \tilde{H}$ 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 $\text{syn}(\tilde{H})$ 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 $\text{syn}(\tilde{H})$.

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 $\tilde{H}$ where $H$ is a hypercode. In particular, we characterize the hypercodes $H$ in the two following cases:

(a) $\text{syn}(\tilde{H})$ is a nil monoid and the alphabet $X$ is finite;
(b) $\text{syn}(\tilde{H})$ 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 $u \in X^*$, $u \not\in H$, $H \cup u$ is not a hypercode. Every hypercode can be embedded in a maximal one. A hypercode is maximal iff $X^* = \tilde{H} \cup H$ [7]. A hypercode is said to be quasi-maximal if $X^* - \{\tilde{H} \cup H\}$ is finite. Since the alphabet $X$ is finite, and since a hypercode over $X$ is always finite, then $\tilde{H}$ is also finite; therefore, $H$ is a quasi-maximal hypercode iff $X^* - \tilde{H}$ 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 $\lg(u) > m$.

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, $\alpha(A) = \{a|a \in X$ and $ras \in A$ 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 $\iff \text{syn}(\tilde{H})$ is a finite nil monoid.

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

(\( \leq \)) Let \( a \in X \). Since \( a(H) = X \), there exist \( r, s \in X^* \) such that \( h_1 = ras \in H \). If \( a \equiv 1(P_{\tilde{H}}) \), then \( h_1 = ras \equiv rs(P_{\tilde{H}}) \). Since \( \tilde{H} \) is a class modulo \( P_{\tilde{H}} \), then \( rs \in \tilde{H} \) and \( h_2 < rs \) for some \( h_2 \in H \). Therefore, \( h_2 < rs < h_1 \) and \( h_2 = rs = h_1 \), a contradiction. Hence \( a \neq 1(P_{\tilde{H}}) \).

Since \( \text{syn}(\tilde{H}) \) is a finite nil monoid, then for every \( a \in X \), there exists \( m > 1 \) such that \( a^m \in \tilde{H} \). Suppose that \( X^* - \{ \tilde{H} \cup \tilde{H} \} = K \) is infinite. Then there exists \( w \in K \) containing at least \( m \) identical letters of the alphabet \( X \), say \( a \). Therefore, \( w = x_1 ax_2 a \cdots ax_m ax_{m+1} \) where \( x_i \in X^* \). Since \( a^m \in \tilde{H} \), then \( a^m < w \) and \( w \in \tilde{H} \), a contradiction. It follows then that \( H \) is quasi-maximal. \( \square \)

Remark that if \( H \) is a quasi-maximal hypercode over a finite alphabet, then \( \text{syn}(\tilde{H}) \) is a nil monoid with a disjunctive \( \mu \)-zero and \( \text{syn}(\tilde{H}) \) 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 \( H \) such that \( \text{syn}(\tilde{H}) \) 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 \( B(\lor, \land) \) is a boolean algebra, then the zero element \( 0 \) of \( B \) (the identity element \( 1 \) of \( B \)) is a disjunctive zero relative to the operation \( \land \) (the operation \( \lor \)). A semilattice or a lattice with a disjunctive zero is not in general a boolean algebra. For example, let \( L = \{0, 1, a_1, a_2, a_3\} \) with \( a_i \land a_j = 0 \) and \( a_i \lor a_j = 1 \) for \( i \neq j \). Clearly, \( 0 \) is a disjunctive element for the operation \( \land \), but \( L \) 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 \( A \) of a monoid \( M \) is said to be cs-prime (or completely semiprime) if \( x^n \in A \), \( n \) a positive integer, implies \( x \in A \). Remark that an ideal \( A \) of a monoid \( M \) is cs-prime iff the quotient monoid \( M/P_A \) is a semilattice with a disjunctive zero.

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

**Proposition 2.** Let \( H \) be a hypercode. Then the following properties are equivalent.

1. \( \tilde{H} \) is cs-prime.
2. \( H \) is power free and completely reflective.
3. \( \text{syn}(\tilde{H}) \) is a semilattice with disjunctive zero.
4. \( \text{syn}(\tilde{H}) \) is a semilattice.

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

$(2) \Rightarrow (1)$. Since $\tilde{H}$ is an ideal, in order to show that $\tilde{H}$ is $cs$-prime, it is sufficient to show that $u^2 \in \tilde{H}$ implies $u \in H$. Suppose that $u \notin \tilde{H}$. Then there exists $v \in H$ such that $v < u^2$ and $v \notin u$. 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 $v_i \neq v_j$ for $i \neq j$. From $v < u^2$, we have $v = v_1 \cdots v_r v_{r+1} \cdots v_m$ 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 $u \in \tilde{H}$, a contradiction.

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

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

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

Remark that in Proposition 2 the semilattice $\text{syn}(\tilde{H})$ 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 $\text{syn}(\tilde{H})$, 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