INFORMATION AND CONTROL 34, 123–129 (1977)

Disjunctive Languages On A Free Monoid*

H. J. SHYR

Department of Mathematics, University of Western Ontario, London, Ontario, Canada

A language A on a free monoid X^* generated by X is called a disjunctive language if the principal congruence determined by A is the identity. In this paper we show that if X contains only one letter then the disjunctive languages are exactly the nonregular languages. We construct some disjunctive languages on X^* with $|X| \geqslant 2$ and show that X^* is a disjoint union of infinitely many disjunctive languages. We also show that the family of disjunctive languages is an ANTI-AFL.

1. INTRODUCTION

Let X be an alphabet and let X^* be the free monoid generated by X. Let $X^+ = X^* \backslash \{\Lambda\}$, where Λ is the empty word. For any $w \in X^*$, we let $\lg(w)$ represent the length of w. In particular, $\lg(\Lambda) = 0$. A language A ovre X is a subset of X^*. For any $A \subset X^*$, the relation P_A defined on X^* by $x, y \in X^*$, $x \equiv y(P_A)$ if and only if ($uxv \in A \Leftrightarrow uyv \in A$ for all $u, v \in X^*$) is a congruence, called the *principal congruence* on X^* determined by A. If X is finite, then the language $A \subset X^*$ is called *regular* if and only if the index of P_A is finite. A language A is said to be *disjunctive* if P_A is the identity. This is equivalent to saying that A is disjunctive if and only if for all $x \neq y \in X^*$, there exists $u, v \in X^*$ such that $uxv \in A$ and $uyv \notin A$ or vice versa. In Section 2, we characterize the disjunctive languages over a one letter alphabet. We give an example to show that a disjunctive language need not have infinite gaps. In Section 3 we construct some disjunctive languages over an alphabet with more than one letter. We also study some of their properties. In particular the set of all primitive words is a disjunctive language. In Section 4 we construct another class of disjunctive languages which are not of the type we constructed in Section 3. Finally, in Section 5 we show that the family of disjunctive languages is an ANTI-AFL.

* This research has been supported by Grant A7877 of the National Research Council of Canada.

ISSN 0019-9958

2. Disjunctive Languages over a One Letter Alphabet

In this section we consider the case when X contains only one letter, i.e., $X = \{a\}$. Most of the time we will use a^* instead of X^*.

The following is a characterization of nondisjunctive languages on a^*.

Proposition 1. *Let $A \subset a^*$. Then the following are equivalent*

(1) *A is not a disjunctive language;*

(2) *there exists $N \geqslant 0$, $m \geqslant 1$ such that $a^N a^r \in A$ if and only if $a^N a^m a^r \in A$ for all $r \geqslant 0$;*

(3) *A is regular.*

Proof. (1) implies (2). Suppose A is not a disjunctive language. Then P_A is not the equality. Let $B \subset a^*$ be a congruence class which is not a singleton set and let $N \geqslant 0$, $m \geqslant 1$ be two integers such that a^N, $a^N a^m \in B$, i.e., $a^N \equiv a^N a^m$ (P_A). Since P_A is a congruence we have $a^N a^r \in A$ if and only if $a^N a^m a^r \in A$ for all $r \geqslant 0$.

(2) implies (3). Suppose there exists $N \geqslant 0$ and $m \geqslant 1$ such that $a^N a^r \in A$ if and only if $a^N a^r a^m \in A$ for all $r \geqslant 0$. Then the index of P_A is at most $N + m$. Hence A is regular.

(3) implies (1). Suppose A is regular. Then P_A is of finite index. This implies that P_A is not the equality. Hence A is not a disjunctive language. #

Proposition 2. *Let $A \subset a^*$. Then A is a disjunctive language if the conditions hold*

(1) *for any $m \geqslant 1$ there exists $r \geqslant 1$ such that $a^{m+r} \in A$;*

(2) *for any $m \geqslant 1$ there exists $n \geqslant 1$ such that $a^{m+n} \notin A$;*

(3) *for any $m \geqslant 1$ and for every $n_1 \geqslant 1$ there exists $s \geqslant n_1$ such that either $a^s\{a, a^2,..., a^m\} \cap A = \varnothing$ or $a^s\{a, a,..., a^m\} \subset A$.*

Proof. Suppose A satisfies the conditions. Then by (1) and (2), $A \neq \varnothing$ and $A \neq a^*$. Let a^r, a^{r+k}, $r \geqslant 0$, $k \geqslant 1$ be two words in a^+. If $a^r \in A$, $a^{r+k} \notin A$ or $a^r \notin A$, $a^{r+k} \in A$, then $a^r \not\equiv a^{r+k}(P_A)$. Suppose on the contrary $\{a^r, a^{r+k}\} \cap A = \varnothing$ or $\{a^r, a^{r+k}\} \subset A$. By (3) there exists $s \geqslant r$ such that $a^s\{a, a^2,..., a^{k+1}\} \cap A = \varnothing$ or $a^s\{a, a^2,..., a^{k+1}\} \subset A$. If $a^s\{a, a^2,..., a^{k+1}\} \cap A = \varnothing$, let m_1 be the smallest positive integer such that $a^{s+(k+1)+m_1} \in A$. We have then $a^{s-r+1}a^r a^{m_1} \notin A$ but $a^{s-r+1}a^{r+k}a^{m_1} \in A$. If $a^s\{a, a^2,..., a^{k+1}\} \subset A$, let m_1 be the smallest positive integer such that $a^{s+k+1+m_1} \notin A$. Then $a^{s-r+1}a^r a^{m_1} \in A$ and $a^{s-r+1}a^{r+k}a^{m_1} \notin A$. In either case we have $a^r \not\equiv a^{r+k}(P_A)$. Hence A is a disjunctive language. #

If a disjunctive language A satisfies the conditions in Proposition 2 then either A or $\tilde{A} = a^* \backslash A$ has arbitrarily large gaps. But there exist disjunctive languages which are not of this type. The following is an example:

Let $C = \{n^2 + 2m \mid 0 \leqslant m < n\}$ and $A = \{a^i \mid i \in C\}$. It is easy to see that A is a disjunctive language and A does not satisfy (3) of Proposition 2.

3. Disjunctive Languages on X^* with $|X| \geqslant 2$

Throughout this section we let X be an alphabet with $|X| \geqslant 2$. A word $x \in X^+$ is said to be *primitive* if $x = f^n$ for some $f \in X^+$ implies $n = 1$. We let Q be the set of all primitive words on X^*. If $f \in X^+ \backslash \{a^+ \mid a \in X\}$ is such that $\lg(f) = p \geqslant 3$, where p is a prime, then $f \in Q$. Hence Q is infinite. It is known (see Lyndon and Schützenberger, 1962) that for any $x \in X^+$, there exists a unique $f \in Q$ and $n \geqslant 1$ such that $x = f^n$.

Let $Q^{(1)} = Q \cup \{\Lambda\}$ and for $i \geqslant 2$ we let $Q^{(i)} = \{f^i \mid f \in Q\}$. Then $X^* = \bigcup_{i=1}^{\infty} Q^{(i)}$ with $Q^{(i)} \cap Q^{(j)} = \varnothing$ if $i \neq j$. In this section we discuss some special type of primitive words and show that $Q^{(i)}$ is a disjunctive language for all $i \geqslant 1$.

Lyndon and Schützenberger (1962) have shown that the equation $a^M = b^N c^P$, where $M \geqslant 2, N \geqslant 2, P \geqslant 2$ has a solution on a free group only when a, b, and c are each a power of a common element. Since every free monoid can be embedded in a free group, the result is true on a free monoid. The following lemma is immediate.

Lemma 3. *Let $f, g \in Q, f \neq g$. Then $f^n g^p \in Q$ for all $n \geqslant 2, p \geqslant 2$.*

This lemma is not true if $p = 1$. For example, let $X = \{a, b\}$, $f = ab$, $g = bababb$. Then $f^2 g = (abab)(bababb) = (ababb)^2 \notin Q$.

Lemma 4. *Let $f, g \in Q$, $f \neq g$, and $n \geqslant 2$. If $fg^n \notin Q$, then $fg^{n+k} \in Q$ for all $k \geqslant 2$.*

Proof. Suppose $fg^n = h^r$, $h \in Q$, $r \geqslant 2$. Then $g \neq h$. By Lemma 3 for all $k \geqslant 2$, $fg^{n+k} = h^r g^k \in Q$. #

A subset $A \subset X^+$ is called a *code* if $x_1 x_2 \cdots x_m = y_1 y_2 \cdots y_n$, $x_i, y_j \in A$, $m \geqslant 1, n \geqslant 1$, implies $m = n$ and $x_i = y_i$, $i = 1, 2, \ldots, n$. A code is therefore a subset of X^+ which generates a free submonoid of X^*.

Lemma 5. *Let X be an alphabet. Then for any $x, y \in X^+$, $\{x, y\}$ is a code if and only if $xy \neq yx$.*

Proof. This is a direct consequence of Lemma 3 (Lyndon and Schützenberger, 1962) and Corollary 3 (Lentin and Schützenberger, 1967). #

Proposition 6. *Let $f \neq g, f, g \in Q$. Then $g^i f^m g^i \in Q$ for all $m \geqslant 2, i \geqslant 1$.*

Proof. Suppose $g^i f^m g^i = h^r$, $h \in Q$, $r \geqslant 2$. If $f \neq h$, then $(g^i f^m)(g^i f^m) = h^r f^m \in Q$, a contradiction. Hence $f = h$ must hold. It follows that either $f = gk$,

$k \in X^+$ or $g = fq$, $q \in X^+$. Now if $f = gk$, then $gk \neq kg$ and hence by Lemma 5 $\{g, k\}$ is a code. We have then

$$g^i gk \cdots gkg^i = gkgk \cdots gk.$$

The fact that $\{g, k\}$ is a code implies that $g = k$, a contradiction. On the other hand, if $g = fq$, $q \in X^+$, then $fq \neq qf$ and $\{f, q\}$ is a code. We have

$$(fq)^i f \cdots f(fq)^i = f \cdots f.$$

Again this implies that $f = q$, a contradiction. Therefore $g^i f^m g^i \in Q$ for all $m \geqslant 2$, $i \geqslant 1$. #

The above result is not true if $m = 1$. For example, let $X = \{a, b\}$. Then $(a^i)(ba^ia^ib)(a^i) = (a^iba^i)(a^iba^i) \notin Q$, where a and ba^ia^ib are primitive words.

Let $A, B \subset X^*$ be two nonempty languages such that $A \cap B = \varnothing$. The pair (A, B) is said to be a *disjunctive pair* if for all $x, y \in X^*$, $x \neq y$, there exist $u, v \in X^*$ such that either $uxv \in A$, $uyv \in B$ or $uyv \in A$, $uxv \in B$. Note that A is a disjunctive language if and only if $(A, X^* \backslash A)$ is a disjunctive pair. If (A, B) is a disjunctive pair, then both A and B are disjunctive languages. Moreover, if (A, B) is a disjunctive pair and $A_1 \supset A$, $B_1 \supset B$, $A_1 \cap B_1 = \varnothing$, then (A_1, B_1) is also a disjunctive pair.

PROPOSITION 7. *Let $X = \{a, b, \ldots\}$ be an alphabet such that $a \neq b$. Then $(Q, Q^{(i)})$ is a disjunctive pair for all $i \geqslant 2$.*

Proof. Given i, $x \neq y$, we must prove there exist u, v such that $uxv \in Q^{(i)}$ and $uyv \in Q$ or vice versa. Without loss of generality, assume that $x \neq \Lambda$ and that x begins with b. Then take $M = 2\max(\lg(x), \lg(y))$, $u = a^M x$, $v = (a^M xx)^{i-1}$. Since $a^M xx \in Q$ by Lemma 3, $uxv = (a^M xx)^i \in Q^{(i)}$.

We prove $uyv = a^M xy(a^M xx)^{i-1} \in Q$, by attempting to find $j \geqq 2$ and w such that $w^j = a^M xy(a^M xx)^{i-1}$. We can easily eliminate $w \in a^*$. Thus $w = a^M bw_1$, for some w_1. Since w begins with exactly M a's, we can infer that $w = a^M xy(a^M xx)^h = (a^M xx)^{h+1}$, for some $h \geqq 0$. But this is impossible since $x \neq y$. So with $j \geqq 2$, no such w can be found.

Remark. It is immediate that $(Q^{(1)}, Q^{(i)})$ is also a disjunctive pair for all $i \geqslant 2$, where $Q^{(1)} = Q \cup \{\Lambda\}$.

We have the following as a corollary,

PROPOSITION 8. *Let X be an alphabet such that $|X| \geqslant 2$. Then each $Q^{(i)}$, $i = 1, 2, \ldots$ is a disjunctive language. Hence X^* is a disjoint union of infinitely many disjunctive languages. Moreover, if $A \supset Q$, $B \supset Q^{(i)}$ for some $i \geqslant 2$ and $A \cap B = \varnothing$, then (A, B) is a disjunctive pair and hence both A and B are disjunctive languages.*

4. Another Disjunctive Language on X^* with $|X| \geqslant 2$

In this section we let $X = \{a, b\}$. We construct disjunctive languages on X^* with $|X| = 2$ which are different from those of $\underset{\sim}{Q}^{(i)}$ in the previous section. In fact the same construction works for the case $|X| \geqslant 3$.

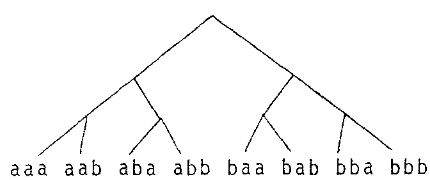
First we define an integer function $f(n)$ which we need in the construction.

$$f(1) = 1,$$
$$f(n) = f(n-1) + n, \qquad \text{if} \quad n \neq 1.$$

We note that $f(n) - f(n-1) = n$ for $n \geqslant 2$.

We let $\leqslant$ be the lexicographic order on X^n, $n \geqslant 1$. Then $(X^+, \leqslant)$ is a partially ordered semigroup.

Let $\#m(S)$ be the mth element in the sequence $S \subset X^n$. For example, $X^3 = \{aaa < aab < aba < \cdots < bba < bbb\}$ and $\#3(X^3) = \{aba\}$.



Let

$$C = \bigcup_{n=1}^{\infty} \{bX^{f(n)}\}$$

and for $k = 1, 2, \ldots$, we let

$$B_k = \bigcup_{n=1}^{\infty} \{\#n(aX^{f(10n+10k)+k})\},$$

and

$$B = \bigcup_{k=1}^{\infty} B_k .$$

Remark. It is easily seen that for all $x \in X^*$ there exist $u, v \in X^+$ such that $uxv \in B$, since if $\#m(X^n) = \{x\}$, $n \geqslant 1$, $1 \leqslant m \leqslant 2^n$, then for all $r \geqslant 1$, $\#m(X^{n+r}) = \{a^r x\}$.

From the construction of C and B we see that

(1) $B \cap C = \varnothing$;

(2) for any $m \geqslant 1$ and $k \geqslant 1$, there exists an integer $j \geqslant k$ such that $(\bigcup_{i=1}^{m} bX^{j+i}) \cap C = \varnothing$; and $bX^{j+m+1} \subset C$;

(3) in the set B, aX^r contains at most one element, $r \geqslant 1$.

PROPOSITION 9. *Let $X = \{a, b\}$. Then the language $A = B \cup C$ is a disjunctive language on X^*.*

Proof. Since $C \subset A$ and $B \cap C = \varnothing$, it follows from (2) that for any $x, y \in X^*$ with $\lg(x) \neq \lg(y)$, $x \not\equiv y(P_A)$. It remains to be shown that $x \not\equiv y(P_A)$ for all $x, y \in X^*$ with $\lg(x) = \lg(y)$. But this is immediate from (3) and the remark following the definition of B. #

5. THE FAMILY OF DISJUNCTIVE LANGUAGES IS ANTI-AFL

Abstract families of languages (AFL) have been studied intensively by Ginsburg, Greibach, Harrison, Spanier, and others The families of regular languages, context-free languages, context-sensitive languages and type-0 languages are examples of AFL. An example of an anti-AFL, where a family of languages $\mathscr{L}$ is called anti-AFL if it is not closed under any of the operations defined in (4) below, is the family of OL-*languages* (see Salomaa, 1973, Theorem 13.1). In this section we show that the family of disjunctive languages is another *anti*-AFL.

In this section we let X be an infinite alphabet and for every i, X_i is a finite alphabet such that $X_i \subset X$.

A mapping h from X_1^* into X_2^* is a *homomorphism* if $h(xy) = h(x)\, h(y)$ for all $x, y \in X_1^*$. If $h(x) = \Lambda$ implies $x = \Lambda$, then h is said to be Λ-*free*. The mapping h^{-1} from subsets of X_2^* into subsets of X_1^* defined by $h^{-1}(Y) = \{x \in X_1^* \mid h(x) \in Y\}$ for all $Y \subset X_2^*$ is called an *inverse homomorphism*.

An abstract family of languages (abbreviated AFL) (see Ginsburg and Greibach, 1969) is a pair $(X, \mathscr{L})$, or $\mathscr{L}$ when X is understood, where

(1) X is an infinite alphabet;

(2) for each A in $\mathscr{L}$ there is a finite set $X_1 \subset X$ such that $A \subset X_1^*$;

(3) $A \neq \varnothing$ for some A;

(4) $\mathscr{L}$ is closed under the operations of union, concatenation, $+$, Λ-free homomorphism, inverse homomorphism and intersection with regular languages.

We now show the following

PROPOSITION 10. *The family of disjunctive languages $\mathscr{D}$ is anti*-AFL.

Proof. (i) Let $X_1 = \{a\} \subset X$ and let $A \neq \varnothing$ be a disjunctive language on X_1^*. Then $\tilde{A} = X_1^* \backslash A$ is also a disjunctive language on X_1^*. But $A \cup \tilde{A} = X_1^*$ is not a disjunctive language. Hence $\mathscr{D}$ is not closed under union.

(ii) Let $X_1 = \{a\} \subset X$ and $A = \{a^{f(n)} \mid n > 1\}$ where $f(n)$ is the integer function defined in the previous section. Then A is disjunctive by Proposition 2 and so $A_1 = \tilde{A}$ is a disjunctive language and $A_1A_1 = a^* = A_1^+$, which is not a disjunctive language. Hence $\mathscr{D}$ is not closed under concatenation and $+$.

(iii) $\mathscr{D}$ is not closed under intersection with regular languages, since the empty set is regular but is not a disjunctive language.

(iv) Let $X_1 = \{a\}$ and $X_2 = \{a, b\}$, where $a \neq b$. Let h be the homomorphism from X_1^* into X_2^* defined by $h(a^n) = (ab)^n$, $n \geqslant 1$. Then h maps the disjunctive language $D = \{a^n \mid n = \text{prime}\} \subset X_1^*$ to $B = h(D) = \{(ab)^n \mid n = \text{prime}\} \subset X_2^*$ which is not in $\mathscr{D}$, since $ab^2a \equiv ab^3a(P_B)$. Hence $\mathscr{D}$ is not closed under Λ-free homomorphism. Now if we let $g: X_2^* \to X_1^*$ such that $g(a) = g(b) = a$. Then $Y = \{a^n \mid n = \text{prime}\}$ is a disjunctive language in X_1^* but $C = g^{-1}(Y) = \{x \in X_2^* \mid \lg(x) = \text{prime}\}$ is not a disjunctive language, since $a \equiv b(P_C)$. Hence $\mathscr{D}$ is not closed under inverse homomorphism. This completes the proof of the proposition. #

Acknowledgment

The author would like to thank the referee for comments and suggesting a shorter proof of Proposition 7.

Received: February 27, 1976; revised: June 11, 1976

References

Ginsburg, S., and Greibach, S. (1969), Abstract families of languages, *in* "Studies in Abstract Families of Languages," Vol. 87 pp. 1–32, Memoirs of the American Mathematical Society.

Lentin, A., and Schützenberger, M. P. (1967), A combinatorial problem in theory of free monoids, *in* "Combinatorial Mathematics and its Applications," pp. 128–144, Proceedings of the Conference Held at the University of North Carolina.

Lyndon, R. C., and Schützenberger, M. P. (1962), On the equation $a^M = b^N c^P$ in a free group, *Michigan Math. J.* **9**, 289–298.

Rabin, M. O., and Scott, D. (1959), Finite automata and their decision problems, *IBM J. Res. Develop.* **3** (2), 114–125.

Rozenberg, G., and Doucet, P. G. (1971), On OL-languages, *Inform. Contr.* **19**, 302–318.

Salomaa, A. (1973), "Formal Languages," Academic Press, New York.