INFORMATION AND CONTROL 9, 620-648 (1966)

Deterministic Context Free Languages*†

SEYMOUR GINSBURG

System Development Corporation, Santa Monica, California

AND

SHEILA GREIBACH

Harvard University, Cambridge, Massachusetts

A number of results about deterministic languages (languages accepted by pushdown automata with no choice of moves) are established. In particular,

(1) each deterministic language is unambiguous.

(2) the complement of each deterministic language is a deterministic language.

(3) numerous operations which preserve deterministic languages (for example, intersection with a regular set) are obtained.

(4) several problems are shown to be recursively unsolvable.

INTRODUCTION

One of the problems associated with a context free language (abbreviated "language") is to find a pushdown automaton (abbreviated "pda") which accepts the language in a relatively efficient manner. If a pda has no choice of moves, then "backtrack" is eliminated and, in a certain sense, the syntactic analysis can be done efficiently. In this paper we study pda with no choice of moves (called "deterministic pda") and the languages (called "deterministic languages") accepted by such pda. It is shown in (Knuth, 1965) that these languages are particularly easy to parse by a so-called left to right translation.

The material is divided into five sections. In the first, after recalling the basic definitions about languages and pda, we formalize the notions

* This work was supported in part by Air Force Cambridge Research Laboratories, Contract AF 19(628)-3418, Project No. 5632, Task No. 563205.

† Presented at the Sixth Annual Symposium on Switching Circuit Theory and Logical Design, Ann Arbor, Michigan, October 7, 1965.

of a deterministic pda and a deterministic language. In Section II we prove the theoretically important facts that if L is a deterministic language, then (i) the complement of L is deterministic, and (ii) L is an unambiguous language.[1] Section III concerns operations which preserve deterministic languages, i.e., functions f such that $f(L)$ is deterministic if L is deterministic. In particular, we prove that intersection with a regular set, as well as the inverse of a generalized sequential machine mapping, preserve deterministic languages. Several operations which do not preserve arbitrary languages but do preserve deterministic languages are exhibited. In addition, a number of common operations which preserve languages, such as word reversal, product, union, and * are shown not to preserve deterministic languages. In Section IV, two results on sequences in deterministic languages are presented. The first yields the fact that $\{a^n b^n / n \geqq 1\} \cup \{a^n b^{2n} / n \geqq 1\}$ is not a deterministic language. The second asserts that a deterministic language contains a sequence if and only if it contains an ultimately periodic sequence, a fact not true for arbitrary languages. The last section is on decision problems. First we show that it is solvable to determine if for an arbitrary deterministic language L and arbitrary regular set R, $L = R$. Then we prove that it is unsolvable to determine whether (i) a language is deterministic, (ii) the union or product of deterministic languages is deterministic, and (iii) L^* is deterministic if L is deterministic.

I. PRELIMINARIES

We shall first review some of the more elementary concepts pertaining to context free languages and pushdown automata. Upon completion of this we shall then introduce the pushdown automata and languages of concern to us.

Definition. A *context free grammar* (abbreviated *grammar*) is a 4-tuple $G = (V, \Sigma, P, \sigma)$ where (i) V is a finite nonempty set; (ii) $\Sigma \subseteq V$ (the set of (*terminal*) *letters*); (iii) P is a finite set of ordered pairs (ξ, w) where ξ is in $V - \Sigma$ and w is in V^*;[2] (iv) σ is in $V - \Sigma$.

[1] These two results have been obtained for a different formulation of deterministic pda concurrently and independently by L. Haines (1965). His arguments differ from ours. (i) has been noted by Fischer (1963), but no proof given. (ii) has also been given by Schutzenberger (1963) for a different formulation of deterministic pda.

[2] For sets of words X and Y, $XY = \{xy/x \text{ in } X, y \text{ in } Y\}$, where xy denotes the concatenation of x and y. XY is called the *product* of X and Y. Let $X^0 = \{\epsilon\}$, where ϵ is the empty word, $X^{i+1} = X^iX$, and $X^* = \bigcup_{i=0}^{\infty} X^i$. Thus, for an arbitrary set E of symbols, E^* is the free semigroup generated by E.

Each element of $V - \Sigma$ is called a *variable.*

Each element (u, v) in P is called a *production* (or *rewriting rule*) and is written $u \to v$.

Notation. Let $G = (V, \Sigma, P, \sigma)$ be a grammar. For words w_1 and w_2 in V^*, write $w_1 \Rightarrow w_2$ if there exist y, u, z, v such that $w_1 = yuz$, $w_2 = yvz$, and $u \to v$ is in P. For words w and x, write $w \Rightarrow^* x$ if either $w = x$ or there exist words $w_0 = w, w_1, \cdots, w_k = x$ such that $w_i \Rightarrow w_{i+1}$ for each i.

A sequence of words $w_0, \cdots, w_k$ such that $w_i \Rightarrow w_{i+1}$ for each i is called a *derivation* or *generation* of w_k (from w_0) and is denoted by

$$w_0 \Rightarrow \cdots \Rightarrow w_k .$$

DEFINITION. $L \subseteq \Sigma^*$ is called a *context free language* (abbreviated *language*) if there exists a grammar $G = (V, \Sigma, P, \sigma)$ such that $L = L(G) = \{w \text{ in } \Sigma^*/\sigma \Rightarrow^* w\}$.

If $L = L(G)$ for the grammar G, then G is said to *generate* L or L is said to be *generated* by G.

We now define recognition devices which are intimately associated with languages.

DEFINITION. A *pushdown automaton* (abbreviated *pda*) is a 7-tuple $M = (K, \Sigma, \Gamma, \delta, Z_0, q_0, F)$ where

(i) K is a nonempty finite set (of *states*).

(ii) Σ is a nonempty finite set (of *inputs*).

(iii) Γ is a finite nonempty set (of *auxiliary* symbols).

(iv) δ is a mapping from $K \times (\Sigma \cup \{\epsilon\}) \times \Gamma$ to the finite subsets of $K \times \Gamma^*$.

(v) Z_0 is an element of Γ.

(vi) q_0 is an K (the *start* state).

(vii) F is a subset of K (the set of *final* states).[3]

Notation. Whenever given a pda M, we shall assume that $M = (K, \Sigma, \Gamma, \delta, Z_0, q_0, F)$.

Notation. Given a pda M let $\vdash_M^*$, or $\vdash^*$ when M is understood, be the relation on $K \times \Sigma^* \times \Gamma^*$ defined as follows. For Z in Γ and x in $\Sigma \cup \{\epsilon\}$ let $(p, xw, \alpha Z) \vdash (q, w, \alpha\gamma)$ if $\delta(p, x, Z)$ contains (q, γ). Let $(p, w, \alpha) \vdash^* (p, w, \alpha)$ for all p, w, α. For α, β in Γ^* and x_i in $\Sigma \cup \{\epsilon\}$

[3] This form of the pda differs slightly from that used by Chomsky (1962, 1963) in that it has a set of final states. This version is selected in order to be homologous with other well-known recognition devices, such as the simple automata, the linear bounded automata etc.

$(1 \leqq i \leqq k)$, write $(p, x_1 \cdots x_k w, \alpha) \vdash^* (q, w, \beta)$ if there exist $p_1 = p, \cdots p_{k+1} = q$ in K and $\alpha_1 = \alpha, \cdots, \alpha_{k+1} = \beta$ in Γ^* such that

$$(p_i, x_i \cdots x_k w, \alpha_i) \vdash (p_{i+1}, x_{i+1} \cdots x_k w, \alpha_{i+1})$$

for $1 \leqq i \leqq$ k.

Speaking informally, a pda has an input tape, a set of states, and a pushdown tape (on which are written auxiliary symbols). The move $(p, xw, \alpha Z) \vdash (q, w, \alpha\gamma)$ means that at state p, with Z the rightmost symbol on the pushdown tape, under x (which is an input symbol or ϵ) the pda goes to state q, writes γ in place of Z, and expends x.

Definition. A word w is *accepted* by a pda M if $(q_0, w, Z_0) \vdash^* (q, \epsilon, \alpha)$ for some q in F and α in Γ^*.[4] Let $T(M)$ denote the set of all words accepted by M.

It is a fact that a set $L \subseteq \Sigma^*$ is a language if and only if $L = T(M)$ for some pda M (Chomsky, 1962).

We now particularize the type of pda with which we shall be concerned. This type, called "deterministic," intuitively is always to have one and only one next move. Such pda are of practical interest. For since they have no nondeterminism, they usually accept or reject words faster than a pda with nondeterminism. (The fact that ϵ serves as an input complicates the comparison.)

Definition. A pda M is said to be *deterministic* if for each q in K and Z in Γ

(a) either $\delta(q, a, Z)$ contains exactly one element for all a in Σ and $\delta(q, \epsilon, Z) = \phi$; or $\delta(q, \epsilon, Z)$ contains exactly one element and $\delta(q, a, Z) = \phi$ for each a in Σ.

(b) if $\delta(q, a, Z_0) \neq \phi$, a in $\Sigma \cup \{\epsilon\}$, then $\delta(q, a, Z_0) = \{(p, Z_0 w)\}$ for some p in K and w in Γ^*.

Condition (a) asserts that either ϵ is applicable, that is, causes the next move, or else all elements in Σ are applicable, but not both. Condition (b) asserts that there is always a non-ϵ word on the pushdown tape (so that a next move is always possible).

Remark. Consider a pda M with the following property: For each q in K and Z in Γ either $\delta(q, a, Z)$ contains at most one element for all a in Σ and $\delta(q, \epsilon, Z) = \phi$; or $\delta(q, \epsilon, Z)$ contains at most one element and

[4] In the version of the pda used by Chomsky (1962), a word w_0 is said to be "accepted" if $(q_0, w_0, Z_0) \vdash (q_1, w_1, \gamma_1) \vdash \cdots \vdash (q_n, w_n, \gamma_n)$ for appropriate q_i, w_i, γ_i, with $q_n = q_0$, $w_n = \gamma_n = \epsilon$, and $q_i \neq q_0$ for each i, $0 < i < n$.

$\delta(q, a, Z) = \phi$ for each a in Σ. It is readily seen that there exists a deterministic pda N such that $T(N) = T(M)$.

DEFINITION. A language L is said to be *deterministic* if $L = T(M)$ for some deterministic pda M.

Many familiar languages are deterministic. Thus, as the reader can easily verify, each regular set[5] L is a deterministic language. Each Dyck language is also deterministic.[6] It will be shown in Theorem 3.1 that $L \cap R$ is deterministic for each deterministic language L and regular set R. Hence the Chomsky-Schutzenberger Normal Form Theorem[7] implies that every language is the homomorphic image of a deterministic language.

It is natural to want to use deterministic languages in syntactic analysis. Oettinger (1961) observed that the Lukasiewicz parenthesis free notation is deterministic, as is the related parenthetic language.

Remark. It is known (Greibach, 1965) that each language is accepted by some pda M in which $\delta(q, \epsilon, Z) = \phi$ for each q in K and Z in Γ. The analogous statement is not true for deterministic languages. Thus there exist deterministic languages accepted by no deterministic pda M in which $\delta(q, \epsilon, Z) = \phi$ for each q in K and Z in Γ. Without giving the details, we note that for $\Sigma = \{a, b, c\}$,

$$L = \{a^ib^ja^i/i, j \geqq 1\} \cup \{a^ib^jcb^ja^i/i, j \geqq 1\}$$

is such a deterministic language.

Let $\Sigma = \{a, b, c\}$. For each n-tuple $w = (w_1, \cdots, w_n)$ of non-ϵ words in $\{a, b\}^*$ let $G(w) = (\Sigma \cup \{\sigma\}, \Sigma, P(w), \sigma)$, where $P(w) = \{\sigma \to a^ibcw_i, \sigma \to a^ib\sigma w_i/1 \leqq i \leqq n\}$. The languages $L(G(w))$ and $[L(G(w))]^R$[8] are deterministic. (These languages are used in showing

[5] An *automaton* is a 5-tuple $(K, \Sigma, \delta, p_1, F)$ where K and Σ are finite nonempty sets (of *states* and *inputs* respectively), δ is a function of $K \times \Sigma$ into K, p_1 is in K, and $F \subseteq K$. The function δ is extended to $K \times \Sigma^*$ by defining $\delta(q, \epsilon) = q$ and $\delta(q, wx) = \delta[\delta(q, w), x]$ for each q in K, w in Σ^*, and x in Σ. A set $R \subseteq \Sigma^*$ is said to be *regular* if there exists an automaton $A = (K, \Sigma, \delta, p_1, F)$ such that $R = T(A)$, where $T(A) = \{w/\delta(p_1, w) \text{ is in } F\}$.

[6] For each $n \geqq 1$ let $\Sigma_n = \{a_i, a_i'/1 \leqq i \leqq n\}$ be a set of $2n$ symbols. Let $G_n = (\Sigma_n \cup \{\sigma\}, \Sigma_n, P_n, \sigma)$, where $P_n = \{\sigma \to \epsilon, \sigma \to \sigma a_i \sigma a_i' \sigma/1 \leqq i \leqq n\}$. Then $L(G_n)$ is called a *Dyck* language (Chomsky and Schutzenberger, 1963).

[7] This result asserts that for every language L, there exists a Dyck language D, a regular set R, and a homomorphism τ such that $\tau(D \cap R) = L$ (Chomsky and Schutzenberger, 1963).

[8] Let A be an abstract set and $\epsilon^R = \epsilon$. For each word $a_1 \cdots a_t$, every a_i in A, let $(a_1 \cdots a_t)^R = a_t \cdots a_1$. For $B \subseteq A^*$ let $B^R = \{w^R/w \text{ in } B\}$.

various properties of languages recursively unsolvable. (Bar Hillel *et al.*, 1961).)

While a deterministic pda has no choice of movement, it may have several chances to accept a particular word since $x_1 \cdots x_k = x_1 \cdots x_k x_{k+1} \cdots x_t$ if $x_i = \epsilon$ for $i \geqq k + 1$. In order to remove the choice factor, we introduce the following notation and concepts.

Notation. Let M be a deterministic pda. Write $(q, w, \alpha) \vdash^{d*} (p, y, \gamma)$ if (i) $(q, w, \alpha) \vdash^{*} (p, y, \gamma)$, and (ii) $\delta(p, \epsilon, Z) = \phi$ if $\gamma = \mu Z$, Z in Γ.

Definition. Let M be a deterministic pda. A word w is said to be *d- accepted* if, for some q in F, $(q_0, w, Z_0) \vdash^{d*} (q, \epsilon, \gamma)$. Let $T_d(M)$ denote the set of all words d-accepted by M.

Intuitively speaking, a word is d-accepted if (i) the entire word is read by the automaton, (ii) the automaton continues operating under ϵ as much as possible, and (iii) the automaton ultimately ends in some final state.

In the next section we shall study the deterministic languages by using the sets $T_d(M)$.

II. UNAMBIGUITY AND COMPLEMENT

In the present section we first prove that the sets $T_d(M)$ coincide with the deterministic languages. Then we show that for each deterministic language L, (i) L is "unambiguous," and (ii) $\Sigma^* - L$ is a deterministic language.

Notation. For each pda M let $M_c = (K, \Sigma, \Gamma, \delta, Z_0, q_0, K - F)$.

Definition. A deterministic pda M is said to be *loop-free* if $T_d(M) \cup T_d(M_c) = \Sigma^*$.

Thus a deterministic pda M is loop-free if and only if for every w in Σ^* there exist q in K and γ in Γ^* such that $(q_0, w, Z_0) \vdash^{d*} (q, \epsilon, \gamma)$.

Since $T_d(M) \cap T_d(M_c) = \phi$ for each deterministic pda M, if M is loop-free such that $T(M) = T_d(M)$ then $T(M)$ and $\Sigma^* - T(M)$ are both deterministic. To prove that $\Sigma^* - L$ is deterministic for an arbitrary deterministic language L, it thus suffices to (i) produce a loop-free deterministic pda M such that $L = T_d(M)$; and (ii) show the existence of a deterministic pda N such that $T_d(M_c) = T(N)$. This we now do.

Lemma 2.1. *If M is a deterministic pda, then $T(M) = T_d(N) = T(N)$ for some loop-free deterministic pda N.*

Proof: For each p, q in K, Z in Γ, and i in $\{0, 1\}$, let $[p, q, Z, i]$ be an abstract symbol. Let

$$D = \{[p, q, Z, i]/i = 0, 1; \delta(q, \epsilon, Z) \neq \phi; \quad (q, \epsilon, Z) \vdash^{d*} (p, \epsilon, \gamma) \text{ for some } \gamma\}.$$

[Intuitively, $i = 1$ in $[p, q, Z, i]$ indicates that since last advancing the input, the pda M has passed through a final state, while $i = 0$ indicates it has not.]

Let e_1 and e_2 be two symbols not in K. Let N be the deterministic pda $(K_N, \Sigma, \Gamma, \delta_N, Z_0, q_0, F_N)$ where $K_N = K \cup \{e_1, e_2\} \cup D$,

$$F_N = F \cup \{e_1\} \cup \{[p, q, Z, 1]/[p, q, Z, 1] \text{ in } D\},$$

and δ_N is defined as follows (q' and Z' denote arbitrary elements of K and Γ respectively):

(1) $\delta_N(e_1, a, Z) = \{(e_2, Z_0)\}$ for all a in Σ.

(2) $\delta_N(e_2, a, Z) = \{(e_2, Z_0)\}$ for all a in Σ.

(3) For a in Σ, $i = 0, 1$, and $\delta(q, a, Z) \neq \phi$;

$$\delta_N(q, a, Z) = \delta_N([q, q', Z', i], a, Z) = \delta(q, a, Z).$$

(4) Suppose $\delta(q, \epsilon, Z) \neq \phi$ and there exist p, γ such that $(q, \epsilon, Z) \vdash^{d*} (p, \epsilon, \gamma)$.

(a) Let $\delta_N([q, q', Z', 1], \epsilon, Z) = \{([p, q, Z, 1], \gamma)\}$.

(b) If there exist p' in F and γ' such that $(q, \epsilon, Z) \vdash^* (p', \epsilon, \gamma')$, let $\delta_N(q, \epsilon, Z) = \delta_N([q, q', Z', 0], \epsilon, Z) = \{([p, q, Z, 1], \gamma)\}$.

(c) If there are no p' in F and γ' such that $(q, \epsilon, Z) \vdash^* (p', \epsilon, \gamma')$ let $\delta_N(q, \epsilon, Z) = \delta_N([q, q', Z', 0], \epsilon, Z) = \{([p, q, Z, 0], \gamma)\}$.

(5) Suppose $\delta(q, \epsilon, Z) \neq \phi$ and there is no p, γ such that $(q, \epsilon, Z) \vdash^{d*} (p, \epsilon, \gamma)$.

(a) Let $\delta_N([q, q', Z', 1], \epsilon, Z) = \{(e_1, Z_0)\}$.

(b) If there exist p' in F and γ' such that $(q, \epsilon, Z) \vdash^* (p', \epsilon, \gamma')$, let $\delta_N(q, \epsilon, Z) = \delta_N([q, q', Z', 0], \epsilon, Z) = \{(e_1, Z_0)\}$.

(c) If there are no p' in F and γ' such that $(q, \epsilon, Z) \vdash^* (p', \epsilon, \gamma')$, let $\delta_N(q, \epsilon, Z) = \delta_N([q, q', Z', 0], \epsilon, Z) = \{(e_2, Z_0)\}$.

To prove the lemma it suffices to show that for each word w in Σ^*,

(6) $(q_0, w, Z_0) \vdash_N^{d*} (q, \epsilon, \gamma)$ for some q in K_N and γ in Γ^*;

(7) w is in $T(M)$ if and only if w is in $T_d(N)$;

and

(8) w is in $T(N)$ if and only if w is in $T_d(N)$.

Let f be the mapping of $K_N - \{e_1, e_2\}$ onto K defined by $f(q) = q$ for q in K and $f([q, q', Z, i]) = q$, $i = 0$ or 1. Let w be an arbitrary word in Σ^*. Two cases arise.

(α) Suppose that $(q_0, w, Z_0) \vdash_M^{d*} (q, \epsilon, \gamma)$. Then there exist $w_0, \cdots, w_{k-1}$, each w_i in $\Sigma \cup \{\epsilon\}$ and $w = w_0 \cdots w_{k-1}$, $\gamma_0 = Z_0$, $\gamma_1, \cdots, \gamma_k = \gamma$, and $q_1, \cdots, q_k = q$ satisfying the following:

$$(q_0, w_0 \cdots w_{k-1}, \gamma_0) \vdash_M^* (q_1, w_1 \cdots w_{k-1}, \gamma_1) \vdash_M^* \cdots \vdash_M^{d*} (q_k, \epsilon, \gamma_k),$$

and for each $i < k$, $\gamma_i = \mu_i Y_i$, Y_i in Γ, such that either

(9) w_i is in Σ and $(q_i, w_i \cdots w_{k-1}, \gamma_i) \vdash_M^* (q_{i+1}, w_{i+1} \cdots w_{k-1}, \gamma_{i+1})$ is realized by $(q_i, w_i, Y_i) \vdash_M (q_i, \epsilon, v_i)$, v_i in Γ^*;

(10) $w_i = \epsilon$ and $(q_i, w_i \cdots w_{k-1}, \gamma_i) \vdash_M^* (q_{i+1}, w_{i+1} \cdots w_{k-1}, \gamma_{i+1})$ is realized by $(q_i, \epsilon, Y_i) \vdash_M^{d*} (q_{i+1}, \epsilon, \epsilon)$; or

(11) $w_i = \epsilon$, and $(q_i, w_i, \gamma_i) \vdash_M^{d*} (q_{i+1}, \epsilon, \gamma_{i+1})$ is realized by $(q_i, \epsilon, Y_i) \vdash_M^{d*} (q_{i+1}, \epsilon, v_{i+1})$, $v_{i+1} \neq \epsilon$.

Then there exist $p_0 = q_0, \cdots, p_k$ in K_N such that $q_i = f(p_i)$ and

$$(p_i, w_i \cdots w_{k-1}, \gamma_i) \vdash_N (p_{i+1}, w_{i+1} \cdots w_{k-1}, \gamma_{i+1})$$

for each i, and $(p_0, w, Z_0) \vdash_N^{d*} (p_k, \epsilon, \gamma_k)$. In fact, if w_i is in Σ, then $p_{i+1} = q_{i+1}$ is in K by (3). If q_i, w_i, γ_i are as in (10) or (11); then by (4), $p_{i+1} = [q_{i+1}, q_i, Y_i, j_{i+1}]$. $j_{i+1} = 1$ if $p_i = [q_i, q_{i-1}, Y_{i-1}, 1]$ or if $(q_i, \epsilon, Y_i) \vdash_M^* (p, \epsilon, \beta)$ for some p in F and some β. Otherwise $j_{i+1} = 0$. Thus (6) holds. From the definition of F and F_N, (7) and (8) hold.

(β) Suppose there are no q and γ such that $(q_0, w, Z_0) \vdash_M^{d*} (q, \epsilon, \gamma)$. Either

(12) there are no q and γ such that $(q_0, \epsilon, Z_0) \vdash_M^{d*} (q, \epsilon, \gamma)$; or

(13) there exist a in Σ, w', w'', γ_1, γ_2, q_1, q_2 such that $w = w'aw''$,

$$(q_0, w, Z_0) \vdash_M^{d*} (q_1, aw'', \gamma_1) \vdash_M (q_2, w'', \gamma_2),$$

and $(q_2, \epsilon, \gamma_2) \vdash_M^{d*} (q, \epsilon, \gamma)$ is false for every q and γ.

Consider (12). By (5b) and (5c), $\delta_N(q_0, \epsilon, Z_0) = \{(e_1, Z_0)\}$ if $(q_0, \epsilon, Z_0) \vdash_M^* (q, \epsilon, \gamma)$ for some q in F, and $\delta_N(q_0, \epsilon, Z_0) = \{(e_2, Z_0)\}$ if $(q_0, \epsilon, Z_0) \vdash_M^* (q, \epsilon, \gamma)$ is false for every q in F and γ. Therefore $(q_0, w, Z_0) \vdash_N^{d*} (e_1, \epsilon, Z_0)$ if $w = \epsilon$ and $(q_0, \epsilon, Z_0) \vdash_M^* (q, \epsilon, \gamma)$ for some q in F; and $(q_0, w, Z_0) \vdash_N^{d*} (e_2, \epsilon, Z_0)$ otherwise. Thus (6), (7), and (8) hold.

Consider (13). Since $(q_0, w', Z_0) \vdash_M^{d*} (q_1, \epsilon, \gamma_1)$, by ($\alpha$) there exists p_1 in K_N such that $f(p_1) = q_1$ and $(q_0, w', Z_0) \vdash_N^{d*} (p_1, \epsilon, \gamma_1)$. Since $(q_1, a, \gamma_1) \vdash_M (q_2, \epsilon, \gamma_2)$, $(p_1, a, \gamma_1) \vdash_N (q_2, \epsilon, \gamma_2)$ by (3). Since $(q_2, \epsilon, \gamma_2) \vdash_M^{d*} (q, \epsilon, \gamma)$ is false for every q and γ; there exist $r \geqq 2$, $q_2, \cdots, q_r, \gamma', Z_2, \cdots, Z_r$, each Z_i in Γ, such that $\gamma_2 = \gamma' Z_r \cdots Z_2$, $(q_i, \epsilon, Z_i) \vdash_M^{d*} (q_{i+1}, \epsilon, \epsilon)$, $2 \leqq i < r$, and there is no q and γ for which

$(q_r, \epsilon, Z_r) \vdash_M^{d*} (q, \epsilon, \gamma)$. By (4) and (5), there exist $p_2 = q_2, \cdots, p_r$ such that for each $2 \leqq i < r$

(14) $\delta_N(p_i, \epsilon, Z) = \{(p_{i+1}, \epsilon)\}$;

(15) $p_{i+1} = [q_{i+1}, q_i, Z_i, j_{i+1}]$, with $j_{i+1} = 1$ if $j_i = 1$ or $(q_i, \epsilon, Z_i) \vdash_M^* (q, \epsilon, \gamma)$ for some q in F, and $j_{i+1} = 0$ otherwise; and

(16) $\delta_N(p_r, \epsilon, Z_r) = \{(e_1, Z_0)\}$ if $p_r = [q_r, q_{r-1}, Z_{r-1}, 1]$ or $(q_r, \epsilon, Z_r) \vdash_M^* (q, \epsilon, \gamma)$ for some q in F; and $\delta_N(p_r, \epsilon, Z_r) = \{(e_2, Z_0)\}$ otherwise. Thus (6) holds. By (15) and (16), (7) and (8) hold.

Thus the lemma is verified by case analysis.

Remarks. (1) Part of Lemma 2.1 is proved in (Schutzenberger, 1963) for a different formulation of deterministic pda.

(2) An examination of the proof of Lemma 2.1 reveals the following: "Let

$$h = \max \{|\gamma| / \delta_N(p, a, Z) = \{(p', \gamma)\} \text{ for some } a \text{ in } \Sigma \cup \{\epsilon\}, p, p', Z, \gamma\}.$$[9]

In scanning a word w, N has (i) exactly $|w|$ moves of the form $\delta_N(p, a, Z)$, a in Σ; (ii) at most $|w| + 1$ moves of the form $\delta_N(p, \epsilon, Z) = \{(p', vY)\}$, v in Γ^*, Y in Γ; (iii) all other moves of the form $\delta_N(p, \epsilon, Z) = \{(p', \epsilon)\}$." Thus there are at most $h(2|w| + 1)$ symbols on the pushdown tape which can be erased. Thus N accepts or rejects w in at most $h(2|w| + 1) + (2|w| + 1) = (h + 1)(2|w| + 1)$ moves.

LEMMA 2.2. *For each deterministic pda M, there exists a deterministic pda N such that $T_d(M) = T_d(N) = T(N)$.*

Proof: Without loss of generality we may assume that q_0 is not in F. (For otherwise, let s_0 be a symbol not in K. Let

$$M_0 = (K \cup \{s_0\}, \Sigma, \Gamma, \delta_{M_0}, Z_0, s_0, F),$$

where $\delta_{M_0}(s_0, \epsilon, Z) = \{(q_0, Z_0)\}$ and $\delta_{M_0}(q, a, Z) = \delta_M(q, a, Z)$ for q in K, a in $\Sigma \cup \{\epsilon\}$, and Z in Γ. Then M_0 is deterministic, $T_d(M_0) = T_d(M)$, and s_0 in not in F.) For each q in K let q' be an abstract element. Let $N = (K_N, \Sigma, \Gamma, \delta_N, Z_0, q_0, F)$, where $K_N = K \cup \{q'/q \text{ in } K\}$ and δ_N is defined as follows: If $\delta(q, \epsilon, Z) = \phi$ let $\delta_N(q', \epsilon, Z) = \{(q, Z)\}$ and $\delta_N(q, a, Z) = \{(p', \alpha)\}$, a in Σ, where $\delta(q, a, Z) = \{(p, \alpha)\}$. If $\delta(q, \epsilon, Z) \neq \phi$ let $\delta_N(q, \epsilon, Z) = \delta_N(q', \epsilon, Z) = \{(p', \alpha)\}$, where $\delta(q, \epsilon, Z) = \{(p, \alpha)\}$. Then N is deterministic and $T_d(M) = T_d(N)$. Furthermore, $(q_0, w, Z_0) \vdash_N^* (q, \epsilon, \alpha)$ for some q in F and α in Γ^* if and only if $(q_0, w, Z_0) \vdash_N^{d*} (q, \epsilon, \alpha)$. Thus $T(N) = T_d(N) = T_d(M)$.

We are now ready for the main results of this section.

[9] For each word γ, $|\gamma|$ denotes the length of γ.

THEOREM 2.1. *If M is a deterministic pda, then $\Sigma^* - T(M)$ is a deterministic language.*

Proof: By Lemma 2.1, $T(M) = T_d(N)$ for some loop-free deterministic pda N. Then $\Sigma^* - T(M) = \Sigma^* - T_d(N) = T_d(N_c)$, with N_c deterministic. By Lemma 2.2, there is a deterministic pda N' such that $T_d(N_c) = T_d(N') = T(N')$. Hence $\Sigma^* - T(M)$ is deterministic.

Let $\Sigma = \{a, b\}$. The languages $\{a^ib^ia^j/i, j \geqq 1\}$ and $\{a^ib^ja^j/i, j \geqq 1\}$ are deterministic but their intersection is not even a language. From Theorem 2.1, it follows that the deterministic languages are closed under complementation but not under intersection, thus not under union.

THEOREM 2.2. *If M is a deterministic pda, then $T_d(M)$ and $\Sigma^* - T_d(M)$ are deterministic languages.*

Proof: By Lemma 2.2, $T_d(M) = T(N)$ for some deterministic pda N. By Theorem 2.1, $\Sigma^* - T_d(N)$ is deterministic.

Finally, we relate deterministic pda's to unambiguous languages.[10]

THEOREM 2.3. *If M is a deterministic pda, then $T(M)$ and $\Sigma^* - T(M)$ are unambiguous languages.*

Proof: It suffices to show that $T(M)$ is unambiguous. By Lemma 2.1 we may assume that $T(M) = T_d(M)$. Let e be a symbol not in K. We first construct a special pda $N = (K \cup \{e\}, \Sigma, \Gamma, \delta_N, Z_0, q_0, F)$ such that $T(M) = \text{Null}\ (N)$, where

$$\text{Null (N)} = \{w/(q_0, w, Z_0) \vdash_N^* (q, \epsilon, \epsilon) \text{ for some } q \text{ in } K \cup \{e\}\}.$$

Let δ_N be defined as follows:

(1) $\delta_N(q, x, Z) = \delta(q, x, Z)$ for q in $K - F$ and x in $\Sigma \cup \{\epsilon\}$.

(2) Suppose q is in F. If $\delta(q, \epsilon, Z) \neq \phi$ let $\delta_N(q, \epsilon, Z) = \delta(q, \epsilon, Z)$. If $\delta(q, \epsilon, Z) = \phi$ let $\delta_N(q, \epsilon, Z) = \{(e, \epsilon)\}$ and $\delta_N(q, x, Z) = \delta(q, x, Z)$ for x in Σ.

(3) $\delta_N(e, \epsilon, Z) = \{(e, \epsilon)\}$.

It is easily seen that Null $(N) = T(M)$.

Note that N need not be deterministic. In fact, both conditions (a) and (b) in the definition of a deterministic pda may be violated.

[10] A derivation $\sigma = w_0 \Rightarrow w_1 \Rightarrow \cdots \Rightarrow w_r$ in $G = (V, \Sigma, P, \sigma)$ is said to be *leftmost* if for each i there exist u_i, ξ_i, v_i, y_i such that $w_i = u_i \xi_i v_i$, $w_{i+1} = u_i y_i v_i$; $\xi_i \rightarrow y_i$ is in P, and u_i is in Σ^*. G is said to be *unambiguous* if each word w in $L(G)$ has at most one (and thus exactly one) leftmost derivation generating it. A language is said to be *unambiguous* if there exists some unambiguous grammar generating it.

For each q, r in $K \cup \{e\}$ and each Z in Γ, let $[r, Z, q]$ be an abstract symbol. (Intuitively, $[r, Z, q]$ is to represent the set of all words α in Σ^* such that $(r, \alpha, Z) \vdash_N^* (q, \epsilon, \epsilon)$.) Let

$$V = \{[r, Z, q]/q, r \text{ in } K \cup \{e\}, Z \text{ in } \Gamma\} \cup \Sigma.$$

For each element p in $K \cup \{e\}$ let $G_p = (V, \Sigma, P, [q_0, Z_0, p])$ where P consists of productions of the following kind:

(4) If $\delta_N(r, a, Z)$ contains $(q, Z_1 \cdots Z_k)$, each Z_i in Γ; then

$$[r, Z, s_1] \to a[q, Z_k, s_k][s_k, Z_{k-1}, s_{k-1}] \cdots [s_2, Z_1, s_1]$$

is in P for every sequence $s_1, \ldots, s_k$ of elements in $K \cup \{e\}$.

(5) If $\delta_N(r, a, Z)$ contains (q, ϵ), then $[r, Z, q] \to a$ is in P.

It is a straightforward matter[11] (using induction) to show that

(6) $[r, Z, q] \Rightarrow^* \alpha$, α in V^*, if and only if either

(6a) $\alpha = wy$ for some w in Σ^* and y in $(V - \Sigma)V^*$, and there exist $p_1, \cdots, p_{m+1} = q$, $U_1, \cdots, U_m$ such that $y = [p_1, U_1, p_2] \cdots [p_m, U_m, p_{m+1}]$ and $(r, w, Z) \vdash_N^* (p_1, \epsilon, U_m \cdots U_1)$; or

(6b) α is in Σ^* and $(r, \alpha, Z) \vdash_N^* (q, \epsilon, \epsilon)$.

Now (6b) implies that

$$\text{Null}(N) = \{\alpha/(q_0, \alpha, Z_0) \vdash_N^* (q, \epsilon, \epsilon) \text{ for some } q \text{ in } K \cup \{e\}\}$$
$$= \bigcup_{q \text{ in } K} L(G_q)$$

is a language. Let σ be an element not in V. Let $G = (V \cup \{\sigma\}, \Sigma, P', \sigma)$, where

$$P' = P \cup \{\sigma \to [q_0, Z_0, p]/p \text{ in } K \cup \{e\}\}.$$

Then Null $(N) = L(G)$. Note that there is a one to one correspondence between the set of leftmost G-derivations of a word w in Null (N) and the set of sequences of moves

(7) $(q_0, w, Z_0) \vdash_N (q_1, w_1, \alpha_1) \vdash_N \cdots \vdash_N (q_{k-1}, w_{k-1}, \alpha_{k-1}) \vdash_N (q_k, \epsilon, \epsilon)$

of the pda N. Since $\delta_N(p, x, Z)$ contains at most one element for each p, x, and Z; it follows that there is one and only one sequence of the form (7) which yields $(q_0, w, Z_0) \vdash_N^* (q, \epsilon, \epsilon)$, and q must be e. Hence there is one and only one leftmost G-derivation of each word in Null (N), i.e., G is unambiguous. Thus $T(M)$ is unambiguous.

[11] The details are given in (Ginsburg, 1966).

From Lemma 2.2 there follows

THEOREM 2.4. *If M is a deterministic pda, then $T_d(M)$ and $\Sigma^* - T_d(M)$ are unambiguous languages.*

III. OPERATIONS

As we have seen, the family of deterministic languages is not closed under union or intersection. We now examine in detail the problem of which operations preserve deterministic languages. In the course of the investigation we shall find that some of the commonplace operations which preserve languages do not preserve deterministic languages.

In (Bar-Hillel *et al.*, 1961) it was shown that intersection with a regular set preserves languages. We now prove that intersection with a regular set preserves deterministic languages.

THEOREM 3.1. *If M is a deterministic pda and R is a regular set, then $T(M) \cap R$ is deterministic.*

Proof: Let $M = (K, \Sigma, \Gamma, \delta, Z_0, q_0, F)$ and $R = T(A)$, where $A = (K_A, \Sigma, \delta_A, p_0, F_A)$ is an automaton. Let

$$M' = (K \times K_A, \Sigma, \Gamma, \delta', Z_0, (q_0, p_0), F \times F_A),$$

where

$$\delta'((q, p), a, Z) = \{((q_1, \delta_A(p, a)), w)\} \quad \text{if} \quad \delta(q, a, Z) = \{(q_1, w)\}, a \text{ in } \Sigma,$$

and

$$\delta'((q, p), \epsilon, Z) = \{((q_1, p), w)\} \quad \text{if} \quad \delta(q, \epsilon, Z) = \{(q_1, w)\}.$$

It is readily seen that M' is deterministic and $T(M) \cap R = T(M')$.

COROLLARY. *If M is a deterministic pda and R is a regular set, then $T(M) \cup R$, $T(M) - R$, and $R - T(M)$ are deterministic languages.*

Proof: Now $T(M) \cup R = \Sigma^* - [(\Sigma^* - T(M)) \cap (\Sigma^* - R)]$, $T(M) - R = T(M) \cap (\Sigma^* - R)$, and $R - T(M) = R \cap (\Sigma^* - T(M))$. The result then follows from Theorems 2.1, 3.1, and the closure of regular sets under intersection and subtraction (Rabin and Scott, 1959).

Another important operation which preserves languages is a gsm mapping[12] (Ginsburg and Rose, 1963). This operation, however, does

[12] A *gsm (generalized sequential machine)* is a 6-tuple $S = (K, \Sigma, \Delta, \delta, \lambda, p_1)$ where K, Σ, Δ are finite nonempty sets (of *states*, *inputs*, and *outputs* respectively), δ is a mapping of $K \times \Sigma$ into K, λ is a mapping of $K \times \Sigma$ into Δ^*, and p_1 is in K. The functions δ and λ are extended to $K \times \Sigma^*$ by defining $\delta(q, \epsilon) = q$, $\lambda(q, \epsilon) = \epsilon$, $\delta(q, wx) = \delta[\delta(q, w), x]$, and $\lambda(q, wx) = \lambda(q, w)\ \lambda[\delta(q, w), x]$ for each q in K, w in Σ^*, and x in Σ. The function S of Σ^* into Δ^* defined by $S(w) = \lambda(p_1, w)$ for each w is called a *gsm mapping*.

not preserve deterministic languages. For let L_1 and L_2 be two deterministic languages such that $L_1 \cup L_2$ is not deterministic. Let c and d be two symbols not in Σ and let $L = cL_1 \cup dL_2$. Obviously L is deterministic. However, if S is the one-state gsm such that $S(c) = S(d) = \epsilon$ and $S(a) = a$ for all a in Σ, then $S(L) = L_1 \cup L_2$ is not deterministic.

In (Ginsburg and Rose, 1963) it was shown that the inverse of a gsm mapping preserves languages. This operation also preserves deterministic languages.

THEOREM 3.2. *If L is a deterministic language and S is a gsm, then*

$$S^{-1}(L) = \{w/S(w) \text{ in } L\}$$

is a deterministic language.

Proof: Let $S = (K_S, \Delta, \Sigma, \delta_S, \lambda_S, p_0)$ and let $L = T(M)$ for some deterministic pda M. Let 0 be a symbol not in Σ. Let $\bar{q}_0$ and $\bar{q}_1$ be two symbols not in K. Let $n = 1 + \max\{|\lambda(p, a)|/p \text{ in } K_S, a \text{ in } \Delta\}$. For any set E and $j \geqq 1$ let $E^{(j)} = E \times \cdots \times E$ (j-times). Now let N be the pda $(K_N, \Delta, \Gamma, \delta_N, Z_0, \bar{q}_0, F_N)$, where

$$K_N = \{\bar{q}_0, \bar{q}_1\} \cup (K \times K_S) \cup (K \times K_S \times H),$$

$H = \bigcup_{j+m \leqq n} (0^{(j)} \times \Sigma^{(m)} \times 0^{(n-(j+m))})$, $F_N = \{\bar{q}_1\} \cup (F \times K_S \times 0^{(n)})$ if ϵ is in L, $F_N = F \times K_S \times 0^{(n)}$ if ϵ is not in L, and δ_N is defined as follows (Z, p, a, q denote arbitrary elements of Γ, K_S, Δ, K respectively):

(1) $\delta_N(\bar{q}_0, \epsilon, Z) = \{(\bar{q}_1, Z)\}$ and $\delta_N(\bar{q}_1, \epsilon, Z) = \{((q_0, p_0), Z)\}$.

(2) (a) If $\lambda(p, a) = \epsilon$ let $\delta_N((q, p), a, Z)$

$$= \{([q, \delta_S(p, a), 0, \cdots, 0], Z)\}.$$

(b) If $\lambda(p, a) = b_1 \cdots b_r$, $1 \leqq r \leqq n$, each b_i in Σ, let

$$\delta_N((q, p), a, Z) = \{([q, \delta_S(p, a), b_1, \cdots, b_r, 0, \cdots, 0], Z)\}.$$

(3) (a) If $\delta(q, \epsilon, Z) = \{(q', w)\}$ let

$$\delta_N([q, p, 0, \cdots, 0], \epsilon, Z) = \{([q', p, 0, \cdots, 0], w)\}.$$

(b) If $\delta(q, \epsilon, Z) = \phi$ let

$$\delta_N([q, p, 0 \cdots, 0], \epsilon, Z) = \{((q, p), Z)\}.$$

(4) Let $s = [q, p, 0^{(t)}, b_1, \cdots, b_r, 0^{(n-(t+r))}]$, with $1 \leqq r \leqq n - 1$, $0 \leqq t \leqq n - 1$, and each b_i in Σ.

(a) If $\delta(q, \epsilon, Z) = \{(q', w)\}$ let

$$\delta_N(s, \epsilon, Z) = \{([q', p, 0^{(t)}, b_1, \cdots, b_r, 0^{(n-(t+r))}], w)\}.$$

(b) If $\delta(q, \epsilon, Z) = \phi$ let

$$\delta_N(s, \epsilon, Z) = \{([q', p, 0^{(t+1)}, b_2, \cdots, b_r, 0^{(n-(t+r))}], w)\},$$

where $\delta(q, b_1, Z) = \{(q', w)\}$.

Clearly N is a deterministic pda. From (1), ϵ is in $T(N)$ if and only if ϵ is in $S^{-1}(L)$. For $w \neq \epsilon$, w is in $S^{-1}(L)$ if and only if there exist $a_1, \cdots, a_k$ in Δ, $u_1, \cdots, u_k$ in Σ^*, $p_1, \cdots, p_k$ in K_S such that

(5) $w = a_1 \cdots a_k$;

(6) $\delta_S(p_{i-1}, a_i) = p_i$ for $1 \leqq i \leqq k$;

(7) $\lambda(p_{i-1}, a_i) = u_i$ for $1 \leqq i \leqq k$; and

(8) $u_1 \cdots u_k$ is in L.

(5)–(8) occur if and only if (5), (6), (7), and

(9) there exist $q_1, \cdots, q_k$ in K, with q_k in F, $\gamma_1, \cdots, \gamma_k$ in $\Gamma^* - \{\epsilon\}$ such that

$$(q_0, u_1 \cdots u_k, Z_0) \vdash_M^* (q_1, u_2 \cdots u_k, \gamma_1) \vdash_M^* \cdots \vdash_M^* (q_k, \epsilon, \gamma_k).$$

From the manner of construction of δ_N; (5), (6), (7), (9) occur if and only if (5), (6), (7), and

(10) $((q_0, p_0), a_1 \cdots a_k, Z_0) \vdash_N^* ([q_1, p_1, 0^{(n)}], a_2 \cdots a_n, \gamma_1] \vdash_N^*$

$$([q_2, p_2, 0^{(n)}], a_3 \cdots a_k, \gamma_2) \vdash_N^* \cdots \vdash_N^* ([q_k, p_k, 0^{(n)}], \epsilon, \gamma_k).$$

Now (5), (6), (7), (10) hold if and only if $w \neq \epsilon$ is in $T(N)$. Thus $T(N) = S^{-1}(L)$.

We shall see later that the product of two deterministic languages is not necessarily a deterministic language. However we do have

THEOREM 3.3. *For L deterministic and R regular, LR is deterministic.*

Proof: Let M be a loop-free deterministic pda such that $L = T(M)$. Without loss of generality we may assume that q_0 is not in F. Let A be an automaton $(K_A, \Sigma, \delta_A, p_0, F_A)$ such that $R = T(A)$. Let N be the deterministic pda $(K \times 2^{K_A}, \Sigma, \Gamma, \delta_N, Z_0, (q_0, \phi), K \times F_N)$, where $2^{K_A} = \{X/X \subseteq K_A\}$, $F_N = \{(q, Y)/q \text{ in } K, Y \subseteq K_A, Y \cap F_A \neq \phi\}$, and δ_N is defined as follows:

(1) If $\delta(q, \epsilon, Z) = \{(q_1, y)\}$ and q_1 is not in F, then $\delta_N((q, Y), \epsilon, Z) = \{((q_1, Y), y)\}$.

(2) If $\delta(q, \epsilon, Z) = \{(q_1, y)\}$ and q_1 in F, then $\delta_N((q, Y), \epsilon, Z) = \{((q_1, Y \cup \{p_0\}), y)\}$.

(3) If $\delta(q, a, Z) = \{(q_1, y)\}$ and q_1 is not in F, then $\delta_N((q, Y), a, Z) = \{((q_1, \delta_A(Y, a)), y)\}$.[13]

[13] For $X \subseteq K_A$, $\delta_A(X, a) = \{\delta_A(x, a)/x \text{ in } X\}$.

(4) if $\delta(q, a, Z) = \{(q_1, y)\}$ and q_1 is in F, then $\delta_N((q, Y), a, Z) = \{((q_1, \delta_A(Y, a) \cup \{p_0\}), y)\}$.

Now w is in LR if and only if $w = w_1w_2$, with w_1 in L and w_2 in R. This is so if and only if

(5) $(q_0, w_1, Z_0) \vdash_M^* (q_1, w_3, \gamma_1) \vdash_M (q, \epsilon, \gamma_2)$ for some q in F, some w_3, and $\delta_A(p_0, w_2)$ in F_A.

Since M is loop-free, (5) is true if and only if

(6) $((q_0, \phi), w_1w_2, Z_0) \vdash_N^* ((q_1, Y_1), w_3w_2, \gamma_1) \vdash_N ((q, \delta_A(Y_1, w_3) \cup \{p_0\}), w_2, \gamma_2) \vdash_N^* ((q_2, Y_2), \epsilon, \gamma_3)$, with q in F, $\delta_A(p_0, w_2)$ in F_A (thus (q_2, Y_2) is in F_N), and $(q, \omega_2, \gamma_2) \vdash_M^* (q_2, \epsilon, \gamma_3)$.

If (6)is true, then w is in $T(N)$. Conversely, if w is in $T(N)$, then there exist w_1, w_2, and w_3 such that $w = w_1w_2$ and (6) holds. Thus LR is deterministic.

We now prove some lemmas involving regular sets.

LEMMA 3.1. *Let M be a deterministic pda. For each q in K let $R_q \subseteq \Gamma^*$ be a regular set. Then there exists a deterministic pda N such that*

$$T(N) = \{w \mid (q_0, w, Z_0) \vdash_M^* (q, \epsilon, \gamma) \text{ for some } q \text{ and some } \gamma \text{ in } R_q\}.$$

Proof: For each q in K let q' and q'' be abstract symbols. Let $K = \{p_1, \cdots, p_n\}$. For each i let $A_i = (K_i, \Gamma, \delta_i, s_i, F_i)$ be an automaton such that $R_{p_i} = T(A_i)$. Let $N = (K_N, \Sigma, \Gamma_N, \delta_N, Z_N, q_0, F_N)$ where $K_N = \{q, q', q''/q \text{ in } K\}$, $\Gamma_N = \Gamma \times K_1 \times \cdots \times K_n$, $Z_N = (Z_0, s_1, \cdots, s_n)$, $F_N = \{q'/q \text{ in } K\}$, and δ_N is defined as follows (t_i, a, Z, Z_i denote arbitrary elements of K_i, $\Sigma \cup \{\epsilon\}$, Γ, and Γ respectively):

(a) $\delta_N(p_i, \epsilon, (Z, t_1, \cdots, t_n))$

$$= \begin{cases} \{(p_i', (Z, t_1, \cdots, t_n))\} \text{ if } \delta_i(t_i, Z) \text{ is in } F_i. \\ \{(p_i'', (Z, t_1, \cdots, t_n))\} \text{ if } \delta_i(t_i, Z) \text{ is in } K - F_i. \end{cases}$$

(b) $\delta_N(p_i', a, (Z, t_1, \cdots, t_n)) = \delta_N(p_i'', a, (Z, t_1, \cdots, t_n))$

$$= \{(p_j, \epsilon)\} \text{ if } \delta(p_i, a, Z) = \{(p_j, \epsilon)\}.$$

(c) $\delta_N(p'_i, a, (Z, t_1, \cdots, t_n)) = \delta_N(p_i'', a, (Z, t_1, \cdots, t_n))$

$$= \{(p_j, (Z_1, t_1, \cdots, t_n)(Z_2, \delta_1(t_1, Z_1), \cdots, \delta_n(t_n, Z_1)) \cdots (Z_k, \delta_1(t_1, Z_1 \cdots Z_{k-1}), \cdots, \delta_n(t_n, Z_1 \cdots Z_{k-1})))\}$$

if $\delta(p_i, a, Z) = \{(p_j, Z_1 \cdots Z_k)\}$.

Obviously N is a deterministic pda. Let g be the homomorphism of

Γ_N^* into Γ^* defined by $g((Z, t_1, \cdots, t_n)) = Z$ for each $(Z, t_1, \cdots, t_n)$ in Γ_N. A straightforward induction shows

(1) if $(q_0, w, Z_N) \vdash_N^* (p, \epsilon, Y_1 \cdots Y_r)$, p in $\{q, q', q''\}$, $r \geqq 1$, each $Y_i = (Z_i, t_{i1}, \cdots, t_{in})$ in Γ_N; then for each i and j, (i) $t_{ij} = \delta_j(s_j, g(Y_1 \cdots Y_{i-1}))$ and (ii) $(q_0, w, Z_0) \vdash_M^* (q, \epsilon, g(Y_1 \cdots Y_r))$.

(2) if $(q_0, w, Z_0) \vdash_M^* (q, \epsilon, \gamma)$; then there exists Y in Γ_N^* such that $g(Y) = \gamma$ and (i) $(q_0, w, Z_N) \vdash_N^* (q', \epsilon, Y)$ if γ is in R_q, (ii) $(q_0, w, Z_N) \vdash_N^* (q'', \epsilon, Y)$ if γ is in $\Gamma^* - R_q$.

From (1) and (2) it follows that

$$T(N) = \{w/(q_0, w, Z_0) \vdash_M^* (q, \epsilon, \gamma) \text{ for some } q \text{ and some } \gamma \text{ in } R_q\}.$$

LEMMA 3.2. *Let $R \subseteq \Sigma^*$ be a regular set and $E_1, \cdots, E_n$ arbitrary subsets of Σ^*. Let $\Delta = \{a_i/i \leqq n\}$ be a set of n elements. Let τ be the substitution of Δ^* into the subsets of Σ^*[14] defined by $\tau(a_i) = E_i$ for each i. Then $U = \{y \text{ in } \Delta^*/\tau(y) \cap R \neq \phi\}$ is regular.*

Proof: Let $R = T(A)$ for some automaton $A = (K, \Sigma, \delta, s_0, F)$. Let B be the automaton $(K_B, \Delta, \delta_B, \{s_0\}, F_B)$ where $K_B = \{X \subseteq K/X \neq \phi\}$, $\delta_B(X, a_i) = \{\delta(q, w)/q \text{ in } X, w \text{ in } E_i\}$, and $F_B = \{X \subseteq K/X \cap F \neq \phi\}$. For any y in Δ^*, $\tau(y) \cap R \neq \phi$ if and only if $\{\delta(s_0, w)/w \text{ in } \tau(y)\} \cap F \neq \phi$. As is easily seen by induction on the length of y, $\delta_B(\{s_0\}, y) = \{\delta(s_0, w)/w \text{ in } \tau(y)\}$ for each y in Δ^*. Thus $T(B) = U$, whence U is regular.

Notation. For M a pda, s in K, and y in Γ^* let

$$L(M, s, y) = \{w \text{ in } \Sigma^*/(s, w, y) \vdash^* (q, \epsilon, \gamma) \quad \text{for some } q \text{ in } F \text{ and } \gamma \text{ in } \Gamma^*\}.$$

LEMMA 3.3. *Let M be a pda, s a state, and V a regular set. Then the set $U(M, s, V) = \{y \text{ in } \Gamma^*/L(M, s, y) \cap V \neq \phi\}$ is regular.*

Proof: Suppose that s is in F and ϵ is in V. Then $U(M, s, V) = \Gamma^*$ and thus is regular. Therefore suppose that either s is not in F or ϵ is not in V. For each (p, Z, q) in $K \times \Gamma \times K$ let $[p, Z, q]$ be an abstract symbol. Let

$$D = \{[p, Z, q]/(p, w, Z) \vdash^* (q, \epsilon, \epsilon) \quad \text{for some } w \text{ in } \Sigma^*\}.$$

For each q in K and Z in Γ let $Z(q)$ be an abstract element if $(q, w, Z) \vdash^* (p, \epsilon, \gamma)$ for some w in Σ^*, γ in Γ^*, and p in F. For each q let Γ_q be the set of all $Z(q)$.

Let s be an element in K and H_s the set of all words $Z(p_0)[p_1, Z_1, p_0]$

[14] Let A and B be abstract sets. For each element a_i in A let $\tau(a_i)$ be a subset of B^*. Let $\tau(\epsilon) = \{\epsilon\}$ and $\tau(x_1 \cdots x_k) = \tau(x_1) \cdots \tau(x_k)$ for each $k \geqq 1$ and x_i in A. Then τ is called a *substitution* (of A^*).

$\cdots [p_m, Z_m, p_{m-1}]$ for $m \geqq 1$, $p_m = s$, $[p_j, Z_j, p_{j-1}]$ in D, and $Z(p_0)$ in Γ_{p_0}. It is easily seen that H_s is regular. Therefore $\Gamma^*\Gamma_s \cup \Gamma^*H_s$ is regular.[15] Let τ be the substitution defined by $\tau(Z) = \{\epsilon\}$,

$$\tau(Z(q)) = \{w^R \text{ in } \Sigma^*/(q, w, Z) \vdash^* (p, \epsilon, \gamma) \text{ for some } \gamma \text{ in } \Gamma^* \text{ and } p \text{ in } F\},$$

and

$$\tau([p, Z, q]) = \{w^R \text{ in } \Sigma^*/(p, w, Z) \vdash^* (q, \epsilon, \epsilon)\}.$$

Since V is regular, so is V^R (Rabin and Scott, 1959). Let $W = \{w/\tau(w) \cap V^R \neq \varphi\}$. By Lemma 3.2, W is regular. Since the intersection of regular sets is regular (Rabin and Scott, 1959), $(\Gamma^*\Gamma_s \cup \Gamma^*H_s) \cap W$ is regular. Let g be the homomorphism defined by $g(Z) = Z$ for each Z in Γ, $g(Z(q)) = Z$ for each $Z(q)$ in Γ_q, and $g([p, Z, q]) = Z$ for each $[p, Z, q]$ in D. Since a homomorphism of a regular set is regular (Bar-Hillel *et al.*, 1961), $g((\Gamma^*\Gamma_s \cup \Gamma^*H_s) \cap W)$ is regular. To complete the proof, it suffices to show that $U(M, s, V) = g((\Gamma^*\Gamma_s \cup \Gamma^*H_s) \cap W)$.

It is readily seen that $g((\Gamma^*\Gamma_s \cup \Gamma^*H_s) \cap W) = g(\Gamma^*\Gamma_s \cap W) \cup g(\Gamma^*H_s \cap W) \subseteq U(M, s, V)$. Thus suppose that α is in $U(M, s, V)$. Then there exists p in F, w in V, and γ in Γ^* such that $(s, w, \alpha) \vdash^* (p, \epsilon, \gamma)$. Therefore there exist $r \geqq 1$, $\alpha = \alpha_0, \alpha_1, \cdots, \alpha_r = \gamma$, $q_1, \cdots, q_r = p$, and $x_1, \cdots, x_r$ in $\Sigma \cup \{\epsilon\}$ such that $w = x_1 \cdots x_r$ and

$$(s, x_1 \cdots x_r, \alpha_0) \vdash (q_1, x_2 \cdots x_r, \alpha_1) \vdash \cdots \vdash (q_r, \epsilon, \alpha_r).$$

Let $n = \min \{|\alpha_i|/1 \leqq i \leqq r\}$ and let k be the smallest integer i such that $|\alpha_i| = n$. Three cases arise.

(a) $k = 0$. Let $\alpha = \alpha'Z$, Z in Γ. Due to the minimality of $|\alpha|$, for each $i \geqq 1$ there exist non-ϵ β_i such that $\alpha_i = \alpha'\beta_i$. Then

$$(s, x_1 \cdots x_r, Z) \vdash (q_1, x_2 \cdots x_r, \beta_1) \vdash \cdots \vdash (q_r, \epsilon, \beta_r).$$

Then $Z(s)$ is in Γ_s and $x_r \cdots x_1$ is in $\tau(\alpha'Z(s)) \cap V^R$. Thus $\alpha'Z(s)$ is in $\Gamma^*\Gamma_s \cap W$, and $\alpha = \alpha'Z$ is in $g(\Gamma^*\Gamma_s \cap W)$.

(b) $k > 0$ and $n \geqq 1$. Then $\alpha_k = \alpha'\beta_k$, where $|\alpha'| = n - 1$. For $i \geqq 0$ there exist non-ϵ β_i such that $\alpha_i = \alpha'\beta_i$. Then $\alpha = \alpha'ZZ_1 \cdots Z_m$, $m \geqq 1$ (since $k > 0$), each Z, Z_i is in Γ; and there exist $p_0, \cdots, p_{m-1}$, $w_0, \cdots, w_m$ such that $w = w_m \cdots w_0$ and

$$(s, w_m \cdots w_0, ZZ_1 \cdots Z_m) \vdash^* (p_{m-1}, w_{m-1} \cdots w_0, ZZ_1 \cdots Z_{m-1}) \vdash^* (p_1, w_1w_0, ZZ_1) \vdash^* (p_0, w_0, Z) \vdash^* (p, \epsilon, \beta_r).$$

[15] The regular sets are closed under union, product, and * (Rabin and Scott, 1959).

Then $Z(p_0)$ is in Γ_{p_0}, w_0^R is in $\tau(Z(p_0))$, and each w_i^R is in $\tau([p_i, Z_i, p_{i-1}])$. Let $\beta' = Z(p_0)[p_1, Z_1, p_0] \cdots [p_m, Z_m, p_{m-1}]$. Then β' is in H_s, $w_0^R \cdots w_m^R$ is in $\tau(\alpha'\beta') \cap V^R$, and $\alpha'\beta'$ is in W. Thus $\alpha'\beta'$ is in $\Gamma^* H_s \cap W$, and $\alpha = g(\alpha'\beta')$ is in $g(\Gamma^* H_s \cap W)$.

(c) $k > 0$ and $n = 0$. Then $\alpha_k = \epsilon$. Thus $k = r$ and $\alpha_r = \epsilon$. If $|\alpha| = 1$ then $\alpha = Z$ is in Γ and $(s, w, Z) \vdash^* (p, \epsilon, \epsilon)$. Then $Z(s)$ is in Γ_s and w^R is in $\tau[Z(s)] \cap V^R$. Thus $Z(s)$ is in W, thus in $\Gamma^*\Gamma_s \cap W$. Hence $\alpha = g(Z(s))$ is in $g(\Gamma^*\Gamma_s \cap W)$. Suppose $|\alpha| > 1$. Then $\alpha = ZZ_1 \cdots Z_m$, $m \geqq 1$, each Z, Z_i in Γ. Then there exist $p_0, \cdots, p_{m-1}, w_m, \cdots, w_0$ such that $w = w_m \cdots w_0$ and

$$(s, w_m \cdots w_0, ZZ_1 \cdots Z_m) \vdash^* (p_{m-1}, w_{m-1} \cdots w_0, ZZ_1 \cdots Z_{m-1})$$
$$\vdash^* \cdots \vdash^* (p_1, w_1w_0, ZZ_1) \vdash^* (p_0, w_0, Z) \vdash^* (p, \epsilon, \epsilon).$$

As in (b), α is in $g(\Gamma^* H_s \cap W)$.

Thus $U(M, s, V) \subseteq g(\Gamma^*\Gamma_s \cap W) \cup g(\Gamma^* H_s \cap W)$, completing the proof.

COROLLARY. *Let M be a pda and V a regular set. Then*

$$\{y/L(M, s, y) \cap V = \phi\} \quad \text{and} \quad \{y/L(M, s, y) \subseteq V\}$$

are regular sets.

Proof: $\{y/L(M, s, y) \cap V = \phi\} = \Gamma^* - U(M, s, V)$ and $\{y/L(M, s, y) \subseteq V\} = \{y/L(M, s, y) \cap (\Sigma^* - V) = \phi\}$.

THEOREM 3.4. *Let $L \subseteq \Sigma^*$ be a deterministic language and c a symbol. Then $f_c(L) = \{u$ in $(\Sigma - \{c\})^*/ucv$ in L for some $v\}$ is a deterministic language.*

Proof: Let $L = T(M)$, where M is a deterministic pda. Since $c\Sigma^*$ is regular,

$$R_q = \{y/L(M, q, y) \cap c\Sigma^* \neq \phi\}$$

is regular for each q in K by Lemma 3.3. By Lemma 3.1, there exists a deterministic pda N such that

$$T(N) = \{u/(q_0, u, Z_0) \vdash_M^* (q, \epsilon, y) \text{ for some } q \text{ in } K \text{ and } y \text{ in } R_q\}.$$

Clearly $T(N) = \{u/ucv$ in L for some $v\}$. Since $f_c(L) = T(N) \cap (\Sigma - \{c\})^*$, $f_c(L)$ is deterministic.

Remark. The mapping f_c preserves languages since there exists a gsm S such that $S(L) = f_c(L)$.

An immediate consequence of Theorems 3.3 and 3.4 is

COROLLARY 1. *$L \subseteq (\Sigma - \{c\})^*$ is deterministic if and only if Lc is deterministic.*

Remark. It is also true that for $L \subseteq (\Sigma - \{c\})^*$, L is deterministic if and only if cL is deterministic. For first suppose that cL is deterministic. Then $cL = T(M)$ for some loop-free deterministic pda M. Let N be the deterministic pda $(K, \Sigma - \{c\}, \Gamma, \delta_N, Z_0, p_0, F)$, where $\delta_N(q, a, Z) = \delta(q, a, Z)$ for all q in K, a in $(\Sigma \cup \{\epsilon\}) - \{c\}$, and Z in Γ, and $(q_0, c, Z_0) \vdash^{*d} (p_0, \epsilon, \gamma)$ for some γ in Γ^*. Then $T(N) = L$. Now suppose that L is deterministic. Then $L = T(M)$ for some deterministic pda $M = (K, \Sigma - \{c\}, \Gamma, \delta, Z_0, q_0, F)$. Let N be the deterministic pda $(K_N, \Sigma \cup \{c\}, \Gamma, \delta_N, Z_0, p_0, F)$, where p_0 and p_1 are two symbols not in K, $K_N = K \cup \{p_0, p_1\}$, and δ_N is defined as follows: $\delta_N(q, a, Z) = \delta(q, a, Z)$ for all q in K, a in $(\Sigma - \{c\}) \cup \{\epsilon\}$, and Z in Γ; $\delta_N(p_0, c, Z_0) = \{(q_0, Z_0)\}$ and $\delta_N(p_0, a, Z) = \{(p_1, Z_0)\}$ for a in $\Sigma - \{c\}$, Z in Γ; $\delta_N(p_1, \epsilon, Z) = \{(p_1, Z)\}$ for Z in Γ; and $\delta_N(q, c, Z) = \{(p_1, Z_0)\}$ for all (q, Z) such that $\delta(q, \epsilon, Z) = \phi$. Then $T(N) = cL$.

COROLLARY 2. *If L is deterministic and R is regular, then*

$$L/R = \{u/uy \quad \text{is in} \quad L \quad \text{for some} \quad y \quad \text{in} \quad R\}$$

is deterministic.

Proof: Let $L \subseteq \Sigma^*$ and c be a symbol not in Σ. Let S be a one-state gsm which maps c into ϵ and a into a, $a \neq c$. By Theorems 3.1 and 3.2,

$$L' = S^{-1}(L) \cap \Sigma^* cR = \{ucv/uv \quad \text{in} \quad L, v \quad \text{in} \quad R\}$$

is deterministic. For f_c as in Theorem 3.4, $L/R = f_c(L')$. Thus L/R is deterministic.

Remark. If L is a language and R is regular, then L/R is a language (Ginsburg and Spencer, 1963).

COROLLARY 3. *If L is deterministic then Init $(L) = \{u/uv$ in L for some v in $\Sigma^*\}$ is deterministic.*

Proof: Init $(L) = L/\Sigma^*$.

Remark. Init (L) is a language if L is a language (Ginsburg and Rose, 1963).

COROLLARY 4. *If L is deterministic and R is regular, then*

$$\text{Div}\,(L, R) = \{u \mid uR \subseteq L\}$$

is deterministic.

Proof: Since $L \subseteq \Sigma^*$ is deterministic, so is $\bar{L} = \Sigma^* - L$. Let c be a symbol not in Σ and S the gsm of Corollary 2. Then

$$L' = S^{-1}(\bar{L}) \cap \Sigma^* cR = \{ucv/v \quad \text{in} \quad R, uv \quad \text{not in} \quad L\}$$

is deterministic. Thus

$$f_c(L') = \{u \quad \text{in} \quad \Sigma^*/uv \quad \text{not in} \quad L \quad \text{for some} \quad v \quad \text{in} \quad R\}$$

is deterministic, where f_c is as in Theorem 3.4. Then

$$\begin{aligned}\Sigma^* - f_c(L') &= \{u/uy \quad \text{is in} \quad L \quad \text{for all} \quad u \quad \text{in} \quad R\} \\ &= \text{Div}\,(L, R)\end{aligned}$$

is deterministic.

Remark. If L is an arbitrary language, Div (L, R) may not be a language. For example, let $\Sigma = \{a, b, c, d, e\}$, $L = \{a^i b^i c^j d/i, j \geqq 1\} \cup \{a^i b^j c^j e/i, j \geqq 1\}$, and $R = \{d, e\}$. Then Div $(L, R) = \{a^i b^i c^i/i \geqq 1\}$, which is not a language.

We now present two operations which preserve deterministic languages but not arbitrary languages.

Notation. For x, y in Σ^* write $x < y$ if $y = xz$ for some z in $\Sigma^* - \{\epsilon\}$. For $L \subseteq \Sigma^*$, let

$$\text{Min}\,(L) = \{y \quad \text{in} \quad L/x \quad \text{in} \quad \Sigma^* - L \quad \text{if} \quad x < y\}$$

and

$$\text{Max}\,(L) = \{y \quad \text{in} \quad L/x \quad \text{in} \quad \Sigma^* - L \quad \text{if} \quad y < x\}.$$

THEOREM 3.5. *If L is a deterministic language, then Min (L) and Max (L) are deterministic languages.*

Proof: Let M be a deterministic pda such that $L = T(M)$.

First consider Min (L). Let s be a symbol not in K and let N be the deterministic pda $(K \cup \{s\}, \Sigma, \Gamma, \delta_N, Z_0, q_0, F)$, where δ_N is defined as follows:

(1) If q is in $K - F$ and a is in $\Sigma \cup \{\epsilon\}$, let $\delta_N(q, a, Z) = \delta(q, a, Z)$.

(2) If q is in F, then $\delta_N(q, a, Z) = \{(s, Z)\}$ for all a in Σ.

(3) $\delta_N(s, a, Z) = \{(s, Z)\}$ for all a in Σ.

A word w is in $T(N)$ if and only if there exist $w_0 = w, \cdots, w_r = \epsilon$, $\alpha_0 = Z_0, \cdots, \alpha_r, q_0, \cdots, q_r$, with q_r in F, such that

(4) $(q_0, w_0, \alpha_0) \vdash_N \cdots \vdash_N (q_r, w_r, \alpha_r)$.

Now (4) occurs if and only if each $q_i (i < r)$ is in $K - F$ and q_r is in F, thus if and only if

(5) $(q_0, w_0, \alpha_0) \vdash_M (q_1, w_1, \alpha_1) \vdash_M \cdots \vdash_M (q_r, w_r, \alpha_r)$

with each $q_i(i < r)$ in $K - F$ and with q_r in F. Since M is deterministic, this is equivalent to $w = w_0$ being in Min $(T(M))$.

Now consider Max (L). For each q in K let $R_q = \{y/L(M, q, y) \cap \Sigma\Sigma^* = \phi\}$. By the corollary to Lemma 3.3, each R_q is regular. By Lemma 3.1, there exists a deterministic pda N such that

$$\begin{aligned} T(N) &= \{w/(q_0, w, Z_0) \vdash_M^* (q, \epsilon, y) \text{ for some } q \text{ in } F \text{ and } y \text{ in } R_q\} \\ &= \{w \text{ in } L/wx \text{ in } \Sigma^* - L \text{ for every } x \neq \epsilon\} \\ &= \text{Max}(L). \end{aligned}$$

Hence the result.

COROLLARY 1. *If Σ contains at least two elements, then $L = \{ww^R/w$ in $\Sigma^*\}$ is not a deterministic language.*

Proof: Clearly L is a language. Suppose L is deterministic. We may assume that Σ contains a and b. Then Min (L) is deterministic. By Theorem 3.1,

$$\begin{aligned} L' &= \text{Min}(L) \cap (ab)(ab)^*(ba)(ba)^*(ab)(ab)^*(ba)(ba)^* \\ &= \{(ab)^n(ba)^m(ab)^m(ba)^n/m, n \geqq 1, m < n\} \end{aligned}$$

is deterministic. It is easy to construct a gsm S such that $S(L') = \{a^n b^m a^m/m, n \geqq 1, m \leqq n\}$. Since L' is a language, so is $S(L')$. But $S(L')$ is not a language (Ginsburg and Spanier, 1964). Therefore L is not deterministic.

Remark. More strongly, $\{ww^R/w$ in $\Sigma^*\}$ is not a finite union of deterministic languages if Σ contains at least two elements. We omit the proof.

COROLLARY 2. *For $\Sigma = \{a, b\}, L = \{a^i b^i a^j/i, j \geqq 1\} \cup \{a^i b^j a^j/i, j \geqq 1\}$ is not deterministic.*

Proof: Clearly L is a language. Since Max $(L) = \{a^i b^j a^j/i, j \geqq 1, i \neq j\}$ is not a language, L is not deterministic.

Remark. The proofs of Corollaries 1 and 2 show that neither Min nor Max preserve languages.

Results about deterministic languages are not always symmetric. (This occurs because a pda scans an input word from left to right.) For example, let L be deterministic and R regular. Then LR is deterministic Theorem 3.3), but RL may not be deterministic even if R is a two-word set. For let $\Sigma = \{a, b, c\}$ and let $L = \{ca^i b^i a^j/i, j \geqq 1\} \cup \{a^i b^j a^j/i, j \geqq 1\}$. Clearly L is deterministic. Let $R = \{c, c^2\}$. Suppose RL is deterministic. Then $RL \cap c^2 a^* b^* a^* = c^2 L_1$ is deterministic, where $L_1 = \{a^i b^j a^j/i, j \geqq 1\}$

$\cup \{a^ib^ia^j/i, j \geqq 1\}$. By applying twice the remark after Corollary 1 of Theorem 3.4, L_1 is deterministic, a contradiction. Thus RL is not deterministic.

There are several other important operations which preserve languages but not deterministic languages.

(i) There exists a deterministic language L such that L^R is not deterministic. For Let $\Sigma = \{a, b, c\}$ and $L = \{c^2a^ib^ja^j/i, j \geqq 1\} \cup \{ca^ib^ia^j/i, j \geqq 1\}$. Clearly L is deterministic. Suppose that L^R is deterministic. Let f_c be as in Theorem 3.4. Then $f_c(L) = \{a^ib^ia^j/i, j \geqq 1\} \cup \{a^ib^ja^j/i, j \geqq 1\}$ is deterministic, a contradiction.

(ii) There exists a deterministic language L such that L^* is not deterministic. For let $\Sigma = \{a, b, c\}$ and L be the deterministic language

$$\{a^ib^ja^j/i, j \geqq 1\} \cup \{ca^ib^ia^j/i, j \geqq 1\} \cup \{c\}.$$

Suppose that L^* is deterministic. Then $L^* \cap c^2a^*b^*aa^* = c^2L_1$ is deterministic, where $L_1 = \{a^ib^ja^j/i, j \geqq 1\} \cup \{a^ib^ia^j/i, j \geqq 1\}$. This is a contradiction.

We now present one other operation, to be used in Section IV, which preserves deterministic languages.

THEOREM 3.6. *For each word w, $Init\ (w) = \{u/uv = w \text{ for some } v \text{ in } \Sigma^*\}$. If L is a deterministic language, then $g(L) = \{w/Init\ (w) \subseteq L\}$ is a deterministic language.*

Proof: Let M be a deterministic pda such that $L = T_d(M)$. Let N be the deterministic pda $(K \cup \{p_0\}, \Sigma, \Gamma, \delta_N, Z_0, q_0, F)$, where p_0 is a symbol not in K and δ_N is defined as follows. Let $\delta_N(q, a, Z) = \{(p_0, Z)\}$ for all a in Σ if q is in $K - F$ and $\delta(q, \epsilon, Z) = \phi$. Let $\delta(p_0, \epsilon, Z) = \{(p_0, Z)\}$ for all Z in Γ. In all other cases let $\delta_N(q, a, Z) = \delta(q, a, Z)$, a in $\Sigma \cup \{\epsilon\}$. Then $g(L) = T_d(N)$, whence $g(L)$ is deterministic.

Remark. If L is an arbitrary language, then $g(L)$ is not necessarily a language. For let $\Sigma = \{a, b, c\}$ and

$$L = \{a^ib^j/0 \leqq j \leqq i, i \geqq 0\} \cup \{a^ib^jc^k/1 \leqq k \leqq j; i, j \geqq 0\}.$$

Then $g(L) = \{a^ib^jc^k/0 \leqq k \leqq j \leqq i\}$, which is not a language.

IV. SEQUENCES

In this section we prove two results about sequences in connection with loop-free deterministic pda. The first result (Lemma 4.1) presents a condition involving a special sequence which is satisfied by every loop-free deterministic pda. Using this condition we then show that for

$\Sigma = \{a, b\}$, the language $L = \{a^n b^n / n \geqq 1\} \cup \{a^n b^{2n} / n \geqq 1\}$ is not deterministic. (We know of no other way to prove L not deterministic. For L is unambiguous, $\Sigma^* - L$ is unambiguous, Min (L) and Max (L) are languages etc.) The second result asserts that if a deterministic language contains a sequence,[16] then it contains an ultimately periodic sequence.[17] (This fact is not true for arbitrary languages (Ginsburg *et al.*, 1965).)

LEMMA 4.1. *Let M be a loop-free deterministic pda and let a be in Σ. Then either*

(i) *there exists $n \geqq 1$ such that for every $m \geqq 1$, if $(q_0, a^m, Z_0) \vdash^* (q, \epsilon, \gamma)$ for some q and γ then $|\gamma| \leqq n$; or*

(ii) *there exist positive integers m, f, words w, y in Γ^*, Z in Γ, and q in K such that for every $h \geqq 0$*

(a) $(q_0, a^{m+hf}, Z_0) \vdash^{d*} (q, \epsilon, wy^h Z)$.

(b) $(q, a^k, wy^h Z) \vdash^* (q', \epsilon, \gamma)$ *implies* $\gamma = wy^h\gamma', \gamma' \neq \epsilon (k \geqq 0)$.

Proof: Assume that (i) does not hold. We first show that for every $r \geqq 1$ there exist $n_r \geqq 1$, Z_r in Γ, q_r, and x_r, $|x_r| \geqq r$, such that $n_{r+1} > n_r$,

(1) $(q_0, a^n r, Z_0) \vdash^{d*} (q_r, \epsilon, x_r Z_r)$, and

(2) for every $s \geqq 0$, $(q_r, a^s, Z_r) \vdash^* (p_s, \epsilon, \gamma_s)$, p_s in K and γ_s in Γ^*.

Thus suppose the contrary. Then there exists an integer r and an infinite number of sequences

(3) $(q_{s0}, u_{s0}, \gamma_{s0}) \vdash (q_{s1}, u_{s1}, \gamma_{s1}) \vdash \cdots \vdash (q_{sk_s}, u_{sk_s}, \gamma_{sk_s})$,

where $q_{s0} = q_0$, $u_{s0} = a^{m(s)}$, $\gamma_{s0} = Z_0$, $u_{sk_s} = \epsilon$, and $|\gamma_{sk_s}| \leqq r$. Since K and Γ are finite, there exist $s \neq t$ such that $(q_{sk_s}, \gamma_{sk_s}) = (q_{tk_t}, \gamma_{tk_t})$. Suppose that $m(s) = m(t)$. By a change in notation if necessary we may assume that $k_s < k_t$. Since M is deterministic, $q_{si} = q_{ti}$, $u_{si} = u_{ti}$, and $\gamma_{si} = \gamma_{ti}$ for $i \leqq k_s$. Then

$$(q_{tk_s}, \epsilon, \gamma_{tk_s}) \vdash \cdots \vdash (q_{tk_t}, \epsilon, \gamma_{tk_t}).$$

Thus $(q_0, a^{m(s)}, Z_0) \vdash^{d*} (q, \epsilon, \gamma)$ for some q and γ is false, contradicting M being loop-free. Suppose that $m(s) \neq m(t)$, say $m(s) < m(t)$. Since M is deterministic, we have $k_s < k_t$, and $q_{si} = q_{ti}$, $u_{si} = u_{ti}$, $\gamma_{si} = \gamma_{ti}$ for $i \leqq k_s$. Then

$$(q_{tk_t}, a^{m(t)-m(s)}, \gamma_{tk_t}) \vdash \cdots \vdash (q_{tk_t}, \epsilon, \gamma_{tk_t}).$$

[16] Let $\{u_i\}$ be an infinite sequence of elements of Σ. A set U of words is said to *contain the sequence* $\{u_i\}$ if U contains each word $u_1 \cdots u_i$. U is said to *contain a sequence* if U contains some sequence $\{u_i\}$.

[17] A sequence $\{u_i\}$ is said to be *ultimately periodic* if there exist positive integers n_0 and p such that $u_{n+p} = u_n$ for all $n \geqq n_0$.

Thus $n = \max\{|\gamma_{ti}|/1 \leqq i \leqq k_t\}$ satisfies (i), a contradiction. Therefore (1) and (2) hold.

For each $r \geqq 1$ let n_r, q_r, and Z_r be as in (1) and (2). Since K and Γ are finite, there exist i and j, $i < j$, such that $Z_i = Z_j$ and $q_i = q_j$. Then

$$(q_0, a^{n_i}, Z_0) \vdash^{d*} (q_i, \epsilon, x_i Z_i) \quad \text{and} \quad (q_0, a^{n_j}, Z_0) \vdash^{d*} (q_i, \epsilon, x_i y Z_i)$$

for some y in Γ^*, with

(4) $(q_i, a^{n_j - n_i}, Z_i) \vdash^{d*} (q_i, \epsilon, yZ_i)$.

Let $m = n_i$ and $f = n_j - n_i$. Then

(5) for each $h \geqq 0$, $(q_0, a^{m+hf}, Z_0) \vdash^{d*} (q_i, a^{hf}, x_i Z_i) \vdash^{d*} (q_i, \epsilon, x_i y^h Z_i)$.

From (4) we get

(6) if $k \geqq 0$, $h \geqq 0$, and $(q_i, a^k, x_i y^h Z_i) \vdash^* (q', \epsilon, \gamma)$; then $\gamma = x_i y^h \gamma'$, $\gamma' \neq \epsilon$.

Thus (ii) holds.

THEOREM 4.1. *For* $\Sigma = \{a, b\}$, $L = \{a^n b^n/n \geqq 1\} \cup \{a^n b^{2n}/n \geqq 1\}$ *is not deterministic.*

Proof: Let $L = T(M)$ for some loop-free deterministic pda. Then either (i) or (ii) of Lemma 4.1 occurs for the symbol a. Suppose alternative (i) occurs. Then for some q in K and γ in Γ^*, there exist an infinite number of positive integers $g(1), \cdots, g(n), \cdots$, such that $(q_0, a^{g(i)}, Z_0) \vdash^* (q, \epsilon, \gamma)$ for each i. Let k be an integer such that $g(k) > g(1)$. Since $a^{g(1)} b^{g(1)}$ is in $T(M)$, there exist q' in F and γ' in Γ^* such that

$$(q_0, a^{g(1)} b^{g(1)}, Z_0) \vdash^* (q, b^{g(1)}, \gamma) \vdash^* (q', \epsilon, \gamma').$$

Then

$$(q_0, a^{g(k)} b^{g(1)}, Z_0) \vdash^* (q, b^{g(1)}, \gamma) \vdash^* (q', \epsilon, \gamma').$$

Since q' is in F, $a^{g(k)} b^{g(1)}$ is in $T(M)$, a contradiction.

Now suppose that alternative (ii) occurs. Then there exist m, f, w, y, Z, and q such that for all $h \geqq 0$

(1) $(q_0, a^{m+hf}, Z_0) \vdash^{d*} (q, \epsilon, wy^h Z)$.

(2) $(q, a^k, wy^h Z) \vdash^* (q', \epsilon, \gamma)$ implies $\gamma = wy^h \gamma'$, $\gamma' \neq \epsilon (k \geqq 0)$.

Since $a^{m+hf} b^{m+hf}$ is in $T(M)$ for each h, from (1) there exist $q_{h0}, \cdots, q_{hk_h}$, $u_{h0}, \cdots, u_{hk_h}$, $\gamma_{h0}, \cdots, \gamma_{hk_h}$, with $q_{h0} = q$, q_{hk_h} in F, $u_{h0} = b^{m+hf}$, $u_{hk_h} = \epsilon$, $\gamma_{h0} = wy^h Z$ such that

$$(q_{h0}, u_{h0}, \gamma_{h0}) \vdash \cdots \vdash (q_{hk_h}, u_{hk_h}, \gamma_{hk_h}).$$

Suppose there is some h such that each $\gamma_{hi} = wv_{hi}$, $v_{hi} \neq \epsilon$. Then

$$(q_0, a^{m+hf+f} b^{m+hf}, Z_0) \vdash^* (q, b^{m+hf}, wy^{h+1} Z) \vdash^* (q_{hk_h}, \epsilon, wyv_{hk_h}),$$

so that $a^{m+hf+f}b^{m+hf}$ is in $T(M)$, a contradiction. Thus for each $h \geqq 0$, there exists a smallest integer $g(h)$ such that $\gamma_{hg(h)} = w$. Since K is finite, there exists a state p and an infinite set J of integers such that $q_{hg(h)} = p$ for each h in J. Let i and $j \geqq m + 3i$ be in J. Since $a^{m+if}b^{2m+2if}$ is in $T(M)$,

$$(q_0, a^{m+if}b^{2m+2if}, Z_0) \vdash^* (p, u_{ig(i)}b^{m+if}, w) \vdash^* (q_{ik_i}, b^{m+if}, \gamma_{ik_i}) \\ \vdash^* (q', \epsilon, \gamma')$$

for some q' in F and γ' in Γ^*. Similarly

$$(q_0, a^{m+jf}b^{2m+2jf}, Z_0) \vdash^* (p, u_{jg(j)}b^{m+jf}, w) \vdash^* (q_{jk_j}, b^{m+jf}, \gamma_{jk_j}) \\ \vdash^* (q'', \epsilon, \gamma'')$$

for some q'' in F and γ'' in Γ^*. Let $\alpha = 2m + if + jf + |u_{jg(j)}| - |u_{ig(i)}|$. Then $\alpha \geqq 2m + if + jf - (m + if) = m + jf \geqq m + mf + 3if > 2m + 2if$ and

$$(q_0, a^{m+if}b^{\alpha}, Z_0) \vdash^* (p, u_{ig(i)}b^{\alpha-(m+if)}, w) \\ = (p, u_{jg(j)}b^{\alpha-(m+if)+|u_{ig(i)}|-|u_{jg(j)}|}, w) \vdash^* (q'', \epsilon, \gamma'').$$

Since q'' is in F, $a^{m+if}b^{\alpha}$ is in $T(M)$, a contradiction. Thus L is not deterministic.

We now turn to the problem of when L contains a sequence or an ultimately periodic sequence.

LEMMA 4.2. *A deterministic language L contains an ultimately periodic sequence if and only if $f(L) = \{w/\mathrm{Init}\,(w) \subseteq L \cup \{\epsilon\}\}$ is infinite, where $\mathrm{Init}\,(w) = \{u/uv = w$ for some v in $\Sigma^*\}$.*

Proof: Obviously $f(L)$ is infinite if L contains an ultimately periodic sequence. Thus suppose that $f(L)$ is infinite. Since L is a deterministic language, so is $L \cup \{\epsilon\}$. By Theorem 3.6, $f(L) = g(L \cup \{\epsilon\})$ is a language. Then (Bar-Hillel *et al.*, 1961) there exist integers p and q with the following property:

($*$) For each word w, $|w| \geqq p$, in $f(L)$ there exist u, v, x, y, z in Σ^* such that $w = uxvyz$, $|xyv| \leqq q$, $xy \neq \epsilon$, and ux^kvy^kz is in $f(L)$ for all $k \geqq 1$.

Since $f(L)$ is infinite, there exists w, $|w| \geqq p$, in $f(L)$. Let u, v, x, y, z be as in ($*$). Suppose $x \neq \epsilon$. From the definition of $f(L)$, Init $(ux^k) \subseteq f(L)$ for all $k \geqq 1$. Then $f(L)$ contains the ultimately periodic sequence

$$a_1, \cdots, a_m, a_{m+1}, \cdots, a_n, a_{m+1}, \cdots, a_n, a_{m+1}, \cdots$$

where $u = a_1 \cdots a_m (m \geqq 0)$ and $x = a_{m+1} \cdots a_n (m < n)$, each a_i in Σ. Suppose that $x = \epsilon$. Then $y \neq \epsilon$ and Init $(uvy^k) \subseteq f(L)$ for all $k \geqq 1$. Thus $f(L)$ contains the ultimately periodic sequence

$$a_1, \cdots, a_m, a_{m+1}, \cdots, a_n, a_{m+1}, \cdots, a_n, a_{m+1}, \cdots$$

where $uv = a_1 \cdots a_m (m \geqq 0)$ and $y = a_{m+1} \cdots a_n (m < n)$, each a_i in Σ.

THEOREM 4.2. *A deterministic language contains a sequence if and only if it contains an ultimately periodic sequence.*

Proof: It suffices to show that a deterministic language L contains an ultimately periodic sequence if it contains a sequence. Let $f(L)$ be as in Lemma 4.2. Since L contains a sequence, $f(L)$ is infinite. Thus L contains an ultimately periodic sequence.

V. DECISION PROBLEMS

We now consider the decidability of various questions. We first present a solvable question, and then some unsolvable ones. We shall use the fact that all constructions given so far are effective. In particular, it follows from Theorem 4.2, Lemma 4.2, and the decidability of a language being infinite (Bar-Hillel *et al.*, 1961) that it is recursively solvable if an arbitrary deterministic language contains a sequence, or an ultimately periodic sequence. These same two questions are recursively unsolvable for arbitrary languages (Ginsburg *et al.*, 1965).

THEOREM 5.1. *It is recursively solvable to determine for an arbitrary deterministic language L and a regular set R whether $L = R$.*

Proof: $L = R$ if and only if $L' = [L \cap (\Sigma^* - R)] \cup [(\Sigma^* - L) \cap R] = \phi$. Since L is deterministic and R is regular, $\Sigma^* - L$ is deterministic and $(\Sigma^* - L) \cap R$ is a language. Thus L' is a language. Then $L = R$ if and only if $L' = \phi$, which is solvable (Bar-Hillel *et al.*, 1961).

Remark. The same problem for an arbitrary language is unsolvable. In fact, for each Σ with at least two elements, it is unsolvable to determine if $L = \Sigma^*$ for an arbitrary language L (Bar-Hillel *et al.*, 1961).

Turning to unsolvable problems we have

THEOREM 5.2. *It is recursively unsolvable whether an arbitrary language over a two letter alphabet is deterministic.*

Proof: In (Bar-Hillel *et al.*, 1961) a language L_1 and a family of languages

$$\{L(x, y)/x, y \text{ } n\text{-tuples of non-}\epsilon \text{ words in } \{a, b\}^*\},$$

both over $\Sigma = \{a, b\}$, were constructed with the following properties:

(1) L_1 and $L(x, y)$ are deterministic.

(2) $L_1 \cap L(x, y) = \phi$ if and only if $L_1 \cap L(x, y)$ is a language.

(3) It is recursively unsolvable to determine for arbitrary $L(x, y)$ whether $L_1 \cap L(x, y) = \phi$.

Let $\bar{L}_1 = \Sigma^* - L_1$ and $\overline{L(x, y)} = \Sigma^* - L(x, y)$. Consider $L = \bar{L}_1 \cup \overline{L(x, y)}$ and $\bar{L} = \Sigma^* - L$. By (1), $\bar{L}_1$ and $\overline{L(x, y)}$ are deterministic. Thus L is a language. However $\bar{L} = L_1 \cap L(x, y)$, which is a language if and only if $\bar{L} = \phi$. Thus L is deterministic if and only if $\bar{L} = \phi$, which is recursively unsolvable.

Some lesser unsolvable problems are summarized in the following theorem.

THEOREM 5.3. *For arbitrary deterministic languages L and L' it is recursively unsolvable to determine whether*

(a) *$L \cup L'$ is deterministic.*

(b) *$L \subseteq L'$.*

(c) *LL' is deterministic.*

(d) *L^* is deterministic.*

Proof: Let L_1, $\bar{L}_1$, $L(x, y)$, and $\overline{L(x, y)}$ be as in Theorem 5.2.

(a) By the proof in Theorem 5.2, $\bar{L}_1 \cup \overline{L(x, y)}$ being deterministic is recursively unsolvable.

(b) $L_1 \subseteq \overline{L(x, y)}$ if and only if $L_1 \cap L(x, y) = \phi$, which is recursively unsolvable.

(c) Let c be a symbol not in Σ, $L = c^*$, and $L' = c\bar{L}_1 \cup \overline{L(x, y)}$. Then $LL' = \overline{L(x, y)} \cup cc^* (\bar{L}_1 \cup \overline{L(x, y)})$ is deterministic if and only if $\bar{L}_1 \cup \overline{L(x, y)}$ is deterministic, which is recursively unsolvable. (For if LL' is deterministic, then $LL' \cap c\, \Sigma^* = c(\bar{L}_1 \cup \overline{L(x, y)})$ is deterministic. By the remark after Corollary 1 of Theorem 3.4, $\bar{L}_1 \cup \overline{L(x, y)})$ is deterministic. If $\bar{L}_1 \cup \overline{L(x, y)}$ is deterministic, then since cc^* is regular, $cc^*(\bar{L}_1 \cup \overline{L(x, y)})$ is deterministic. From this it readily follows that LL' is deterministic.)

(d) Let c and d be two symbols not in Σ. Then

$$L = \{c\} \cup c\bar{L}_1 d \cup \overline{L(x, y)}\, d$$

is deterministic. Then

$$L^* = [\overline{L(x, y)}\, dc^* \cup cc^*(\bar{L}_1 \cup \overline{L(x, y)})\, dc^* \cup cc^*]^*.$$

If L^* is deterministic, then $L^* \cap c\Sigma^* d = c(\bar{L}_1 \cup \overline{L(x, y)})\, d$ is deterministic, whence $\bar{L}_1 \cup \overline{L(x, y)}$ is deterministic. If $\bar{L}_1 \cup \overline{L(x, y)}$ is deterministic, then $\overline{L(x, y)} = L_1$ and

$$L^* = (L_1 dc^* \cup cc^* \Sigma^* dc^* \cup cc^*)^*.$$

In this case it is a straightforward matter to construct a deterministic pda N such that $T(N) = L^*$. Thus L^* is deterministic if and only if $\bar{L}_1 \cup \overline{L(x, y)}$ is deterministic, which is recursively unsolvable.

Remark. By a suitable recoding, the unsolvability of problems (a)–(d) in Theorem 5.3 can be extended to the case where Σ contains two elements.

In conclusion we mention the following open question. Is it recursively unsolvable to determine if $L_1 = L_2$ for arbitrary deterministic languages L_1 and L_2?

RECEIVED: June 14, 1965

REFERENCES

BAR-HILLEL Y., PERLES M., AND SHAMIR, E. (1961), On formal properties of simple phrase structure grammars. *Z. Phonetik, Sprachwissensch. Kommunikationsforsch.* **14**, 143–172.

CHOMSKY, N. (1962), "Context-Free Grammars and Pushdown Storage," Quarterly Progressive Report No. 65, Research Laboratory of Electronics, Massachusetts Institute of Technology.

CHOMSKY, N. (1963), Formal properties of grammars. *In* "Handbook of Mathematical Psychology," D. Luce, R. Bush, and E. Galanter, eds., Vol. 2. Wiley, New York.

CHOMSKY, N., AND SCHUTZENBERGER, M. P. (1963), The algebraic theory of context-free languages. *In* "Computer Programming and Formal Systems," P. Braffort and D. Hirschberg, eds., pp. 118–161. North-Holland, Amsterdam.

FISCHER, P. C. (1963), On computability by certain classes of restricted Turing machines. *Proc. Fourth Ann. Symp. Switching Circuit Theory and Logical Design, September 1963*, pp. 23–31.

GINSBURG, S. (1966), "The Mathematical Theory of Context Free Languages." McGraw-Hill, New York.

GINSBURG, S., HIBBARD, T. N., AND ULLIAN, J. S. (1965), Sequences in context free languages. *Illinois J. Math.* **9**, 321–337.

GINSBURG, S., AND ROSE, G. F. (1963), Operations which preserve definability in languages. *J. Assoc. for Comput. Mach.* **10**, 175–195.

GINSBURG, S., AND SPANIER, E. H. (1963), Quotients of context free languages. *J. Assoc. Comput. Mach.* **10**, 487–492.

GINSBURG, S., AND SPANIER, E. H. (1964), Bounded ALGOL-like languages. *Trans. Am. Math. Soc.* **113**, 333–368.

GREIBACH, S. A. (1965), A new normal-form theorem for context-free phrase structure grammars. *J. Assoc. Comput. Mach.* **12**, 42–52.

HAINES, L. G. (1965), "Generation and Recognition of Formal Languages." Ph.D. Thesis, Massachusetts Institute of Technology.

Knuth, D. E. (1965), On the translation of languages from left to right. *Inform. Control* 8, 607–639.

Oettinger, A. G. (1961), Automatic syntactic analysis and the pushdown store. In "Structure of Language and Its Mathematical Aspects." *Proc. Symp. Appl. Math.* 12, (Am. Math. Soc.).

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

Schutzenberger, M. P. (1963), Context-free languages and pushdown automata. *Inform. Control*, 6, 246–264.

Statement of ownership, management and circulation required by the Act of October 23, 1962, Section 4369, Title 39, United States Code; of Information and Control

Published Bimonthly at Mt. Royal and Guilford Avenues, Baltimore, Maryland 21202 for Academic Press Inc., 111 Fifth Avenue, New York, N. Y. 10003.

Editors: Leon Brillouin, New York, N. Y.; Colin Cherry, Imperial College, London, England; Murray Eden, Massachusetts Institute of Technology, Cambridge, Mass. 02139; Peter Elias, Massachusetts Institute of Technology, Cambridge, Mass. 02139

Owned by Walter J. Johnson, Thekla Johnson, and Kurt Jacoby, 111 Fifth Avenue, New York, N. Y., 10003

Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages or other securities: none. Paragraphs 2 and 3 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting, also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner. Names and addresses of individuals who are stockholders of a corporation which itself is a stockholder or holder of bonds, mortgages or other securities of the publishing corporation have been included in paragraphs 2 and 3 when the interests of such individuals are equivalent to 1 percent or more of the total amount of the stock or securities of the publishing corporation. Total no. copies printed, average no. copies each issue during preceding 12 months: 1700; single issue nearest to filing date: 1700. Paid circulation (a) to term subscribers by mail, carrier delivery or by other means, average no. copies each issue during preceding 12 months: 1270; single issue nearest to filing date: 1294. (b) Sales through agents, news dealers, or otherwise, average no. copies each issue during preceding 12 months: none; single issue nearest to filing date: none. Free distribution by mail, carrier delivery, or by other means, average no. copies each issue during preceding 12 months: 51; single issue nearest to filing date: 52. Total no. of copies distributed, average no. copies each issue during preceding 12 months: 1321; single issue nearest to filing date: 1346.

(Signed) J. B. Chertok, Vice President