N72-29161

ARBITRARY GRAMMARS GENERATING CONTEXT-FREE LANGUAGES*

by

Brenda S. Baker†

11-72

Center for Research in Computing Technology

Harvard University

Cambridge, Massachusetts 02138

ARBITRARY GRAMMARS GENERATING CONTEXT-FREE LANGUAGES*

by

Brenda S. Baker†

11-72

Center for Research in Computing Technology
Harvard University
Cambridge, Massachusetts 02138

*This research has been supported in part by the National Aeronautics and Space Administration under Grant NGR-22-007-176.

†The author is an NSF Pre-Doctoral Fellow at Harvard University.

Abstract

If G is a grammar such that in each non-context-free rule of G, the right side contains a string of terminals longer than any terminal string appearing between two nonterminals in the left side, then the language generated by G is context-free. Six previous results follow as simple corollaries of this theorem.

Introduction

It is well known that the family of languages generated by arbitrary grammars is precisely the family of recursively enumerable sets, and that grammars with only context-free rules generate only context-free languages, a proper subset of the family of recursive sets. However, the way in which rules containing context interact to generate non-context-free languages is not well understood. Previous investigations of the mechanism of context have shown that certain constraints on the use of context force the language generated to be context-free. In this paper it is shown that grammars obeying a particular weak restriction on the form of rules generate only context-free languages, and that a number of previous results follow as simple corollaries of the theorem presented here. The restriction is that in each non-context-free rule, the right side must contain a string of terminals longer than any terminal string appearing between nonterminals in the left side of the rule. This restriction is of interest, not only for its own sake, but because it shows that the previous results all deal with aspects of the same mechanism.

The earlier results involve restrictions on the form of grammar rules and on the mechanism of derivation. Intuitively, the effect of restrictions on either grammar rules or derivations is to limit the capacity of a derivation to coordinate various segments of the sentence it generates. "Sufficient" limitation of this capacity forces segments of a string to be derived independently, essentially in a context-free manner.

Several of the previous results depend on the generation of terminal substrings, which act as barriers to further transmission of information from one side of the substring to the other. Ginsburg and Greibach (1966) show that if no terminal letters can be used as context, and every rule generates at least one terminal letter, then the language generated is context-free. Book (1972) shows that a grammar generates a context-free language if the left side of every rule contains only one nonterminal, with terminal strings as the only context. Book (1972) also shows that if in every rule of a type 0 grammar, the left context is a string of terminal symbols at least as long as the right context, then the language generated is context-free. In each of the above results, terminal strings interfere with the passing of information between nonterminals. Since terminal strings are "fixed", that is, they cannot change once they are generated, terminal strings do not store much information which can be used as context. At the same time, when a sufficiently long terminal string is generated, no single rule can span the string to pass information between the segments on each side of the string; thus the derivation must proceed independently in the two segments.

A related result is that of Hibbard (1966). It states that a language is context-free if it is generated by a grammar which has a partial ordering $<$ on its symbols, such that in every rule of the grammar every symbol on the left side is "less than" some symbol on the right. Here, successive application of rules gradually establishes a "barrier" across which no information can be passed.

Two somewhat different results are the theorems of Matthews (1963, 1964, 1967). They state that for a grammar, the set of terminal strings generated by left-to-right derivations is context-free, and the set of terminal strings generated by "two-way" derivations is context-free.

As a consequence of the main theorem in this paper, we find that in all of the above results, the same blocking mechanism forces the language generated by a restricted grammar to be context-free. Thus, the six theorems cited above do not isolate six distinct aspects of context, but instead are all based on the same mechanism.

Section 1

This section presents the principal notation used in the paper. For basic results concerning grammars, languages, and gsm mappings, see Ginsburg (1966).

The length of a string w will be denoted by $|w|$. The empty string will be denoted by e.

A grammar is a quadruple $G = (V,\Sigma,R,X)$ where V is a finite set of symbols, $\Sigma \subset V$ is the set of terminal symbols, $X \in V - \Sigma$ is the starting symbol, and R is a finite set of rewriting rules of the form $\alpha_0 Y_1 \alpha_1 \ldots \alpha_{n-1} Y_n \alpha_n \to \alpha_0 w_1 \alpha_1 \ldots \alpha_{n-1} w_n \alpha_n$ with $n > 0$ and each $\alpha_i \in \Sigma^*$, $Y_i \in V - \Sigma$, and $w_i \in V^*$. Thus, in a grammar terminal symbols are not rewritten. Define the relation $\Longrightarrow$ on V^* such that for any $\sigma \to \rho \in R$, and $\alpha,\beta \in V^*$, $\alpha\sigma\beta \Longrightarrow \alpha\rho\beta$ ($\alpha \to \rho$ transforms $\alpha\sigma\beta$). A derivation in G is a sequence $\theta_0,\theta_1,\ldots,\theta_n \in V^*$ such that for $i = 1,2,\ldots,n$, $\theta_{i-1} \Longrightarrow \theta_i$; write $\theta_0 \overset{n}{\Longrightarrow} \theta_n$. The transitive reflexive closure of $\Longrightarrow$ is $\overset{*}{\Longrightarrow}$. The language generated by G is $L(G) = \{w \in \Sigma^* \mid X \overset{*}{\Longrightarrow} w\}$.

A grammar $G = (V,\Sigma,R,X)$ is type 0 if each rule in R is of the form $\alpha Z\beta \to \alpha\gamma\beta$ with $\alpha,\beta,\gamma \in V^*$ and $Z \in V - \Sigma$. Thus, only one nonterminal is rewritten in each rule.

A rewriting rule $\alpha \to \beta$ is context-free if $\alpha \in V - \Sigma$. A grammar $G = (V,\Sigma,R,X)$ is context-free if every rule of R is context-free. A language L is context-free if $L = L(G)$ for some context-free grammar G.

It will be sufficient to give an informal description of gsm's. A generalized sequential machine (gsm) M is a deterministic finite-state machine with output; at each step M outputs a string (which may be the empty string) determined by the previous state and the input symbol read. M has an initial state but no final states. For a string w, $M(w)$ is the string output by M upon reading input w. For a language L, $M(L) = \{M(w) \mid w \in L\}$. It is well known that the family of context-free languages is closed under gsm mappings.

Section 2

In this section we define "terminal-bounded" grammars and prove the main theorem, which states that the language generated by a terminal-bounded grammar is context-free.

Definition 1. Let $G = (V,\Sigma,R,X)$ be a grammar. G is terminal-bounded if each rule in R is of the form

$$\alpha_0 Y_1 \alpha_1 Y_2 \ldots \alpha_{n-1} Y_n \alpha_n \rightarrow \beta_0 Z_1 \beta_1 \ldots \beta_{m-1} Z_m \beta_m$$

where each $\alpha_i, \beta_i \in \Sigma^*$, each $Y_i, Z_i \in V - \Sigma$, and either $n = 1$ or for some j, $0 \leq j \leq m$, and all $k = 1,2,\ldots,n-1$, $|\beta_j| > |\alpha_k|$.

Thus in each non-context-free rule of a terminal-bounded grammar, there must be some terminal string in the right side which is strictly longer than all terminal strings which appear between nonterminals in the left side. The right-hand side need not contain a terminal string strictly longer than a terminal string which appears at the leftmost or rightmost end of the left side of the rule. During a derivation, repeated application of rules involving non-terminal context gradually builds up terminal strings until there is a terminal string longer than any which can appear between two non-terminals in the left side of any rule; a "block" has been set up to prevent any passage of information across that point. Enough blocks are created during each derivation to force the language to be context-free.

<u>Theorem</u>. If G is a terminal-bounded grammar, then $L(G)$ is context-free.

To prove the Theorem, we will use a measure of "non-context-freeness" of a grammar:

<u>Definition 2</u>. Define $N(G) = \sum_{\alpha \to \beta \in R} (|\alpha| - 1)$.

Thus $N(G)$ measures the excess of the number of symbols in the left side of rules in G over the number which would be present in a context-free grammar with the same number of rules. Note that $N(G) = 0$ if and only if G is context-free.

The Theorem will follow from the following Lemma, proved subsequently.

<u>Lemma</u>. Let $G = (V, \Sigma, R, X)$ be a terminal-bounded grammar with $N(G) > 0$. Then there exist a terminal-bounded grammar G_1, a regular set S, and a gsm M such that $L(G) = M(L(G_1) \cap S)$ and $N(G_1) < N(G)$.

Proof of the Theorem. Given a terminal-bounded grammar G, repeated application of the Lemma produces for some $n \leq N(G)$ a series of grammars $G_0 = G$, $G_1,\ldots,G_n$, a series of regular sets $S_1,\ldots,S_n$ and a series of gsm's $M_1,M_2,\ldots,M_n$ such that for $i = 0,1,\ldots,n-1$, $L(G_i) = M_{i+1}(S_{i+1} \cap L(G_{i+1}))$ and $N(G_n) = 0$. Since $L(G_n)$ is therefore context-free, and the family of context-free languages is closed under gsm mappings and intersection with regular sets, it follows that $L(G_i)$ is context-free for $i = 0,1,\ldots,n$. Thus $L(G) = L(G_0)$ is context-free. $\square$

It remains to prove the Lemma. For the proof of the Lemma, we need to define three constants obtained from G.

Definition 3. Define $L_G = \max\{|\alpha| \mid \alpha\gamma \to \beta \in R \text{ and } \alpha \in \Sigma^*\}$,

$R_G = \max\{|\alpha| \mid \gamma\alpha \to \beta \in R \text{ and } \alpha \in \Sigma^*\}$,

and $B_G = \max(\{0\} \cup \{|\alpha| \mid \gamma_1 Y_1 \alpha Y_2 \gamma_2 \to \beta \in R,\ \alpha \in \Sigma^*,\ Y_1,Y_2 \in V - \Sigma,$ and $\gamma_1,\gamma_2 \in V^*\})$.

Thus, B_G is the maximum length of any terminal string appearing between nonterminals in the left side of a rule of G. The maximum lengths of strings which can appear at the left or right ends of the left side of a rule of G are L_G and R_G, respectively.

<u>Proof of the Lemma</u>. To prove the Lemma, two cases must be considered:

Case 1. $L_G > B_G$ or $R_G > B_G$.

Case 2. $L_G \leq B_G$ and $R_G \leq B_G$.

To obtain a G_1 with $N(G_1) < N(G)$ for Case 1, we will take a rule of G which has in its left side a terminal string too long to appear between nonterminals in any rule, and shorten the terminal context of this rule. In the second case, context rules which generate terminal strings too long to be used between nonterminals in the left side of any rule will be split into two rules which partition the context.

<u>Case 1</u>. $L_G > B_G$ or $R_G > B_G$.

We will assume without loss of generality that $R_G \geq L_G$ and $R_G > B_G$ (for $L_G > R_G$, the proof constructs a G_1 symmetric to the one given here). The construction of G_1 is based on the observation that no single rule of G can span an entire string AwB where w is a terminal string of length R_G and A and B are any symbols. Thus, if the initial portion of a derivation generates a sentential form with substring w, then any later step of the derivation must apply a rule either to the left of w independent of what occurs to the right of w, or to the right of w independent of what occurs to the left of w.

The objective in the construction of G_1 is to replace a single rule whose left side contains a terminal string of length R_G by a rule with shorter terminal context. But eliminating the context constraint will result in derivations with applications of the new rule when the old one is not applicable. Accordingly, a new symbol, $\$$, will be used to mark applications of the new rule so that "bad" derivations may be screened out.

Choose a rule $\gamma w \to \beta w$ from R with $w \in \Sigma^*$ and $|w| = R_G$. Let $\$$ be a new symbol, and let $V_1 = V \cup \{\$\}$. Replace the rule $\gamma w \to \beta w$ of R by a new rule $\gamma \to \beta w\$$ to obtain a new set of rules $R_1 = (R \cup \{\gamma \to \beta w\$\}) - \{\gamma w \to \beta w\}$.

Define a new grammar $G_1 = (V_1, \Sigma \cup \{\$\}, R_1, X)$. Clearly G_1 is terminal-bounded. Since $|w| = R_G > 0$, the left side of $\gamma \to \beta w\$$ is shorter than the left side of $\gamma w \to \beta w$ which it replaced, and $N(G_1) < N(G)$.

Now, in G, if the rule $\gamma w \to \beta w$ is applied to a string $\sigma\gamma w\rho$, then the result is $\sigma\beta w\rho$ and the substring w is available for use as context by either β or ρ. Replacing $\gamma w \to \beta w$ by the rule $\gamma \to \beta w\$$ in G_1 ensures that if $\gamma \to \beta w\$$ is applied to a string $\sigma\gamma w\rho$, then in the resulting string $\sigma\beta w\$w\rho$, a substring w is also available for use as context by either β or ρ. Thus, if we can restrict our attention to derivations of G_1 in which the rule $\gamma \to \beta w\$$ is applied only in the presence of right context w, and somehow erase the "extra" occurrences of $w\$$ generated by application of the new rule, we will obtain precisely the set $L(G)$.

Let $S = (V \cup \{\$w\})^*$. Clearly S is a regular set. Intersecting $L(G_1)$ with S selects those words of $L(G_1)$ generated by derivations in which the rule $\gamma \to \beta w\$$ is used only when the string w is at some time generated immediately to the right.

The extra substrings $\$w$ generated in G_1 and used by S to check context will be erased by a gsm M which acts as follows on input from $V_1{}^*$. In its initial mode of operation, M outputs each symbol it reads. But when M reads the symbol $\$$, it enters an erasing mode in which it erases a string of the form $\$w$, thereafter returning to the initial mode of operation. If $\$$ is not followed by w, M's action is undefined. (Note that for any string in S, $\$$ will always be followed by w.)

We claim that $M(L(G_1) \cap S) = L(G)$. We will merely outline the proof, since the omitted details can be added in a straightforward manner.

Claim.1. $L(G) \subseteq M(L(G_1) \cap S)$.

Sketch of the proof. It is sufficient to show for any $y \in V^*$ that if $X \overset{n}{\Longrightarrow} y$ in G, then there exists a $z \in S \cap L(G_1)$ such that $M(z) = y$. The proof of this statement is by an induction on n, based on the following argument. For a string z, ~~if~~ $M(z) = y$ then z is simply y with a string $w(\$w)^i$ (for some i) substituted for each occurrence of w in y. Suppose that a rule $u \to v$ is applied to y in G. If there is no overlap between any occurrence of w and the substring u of y to which the rule is applied, then there is a corresponding substring u in z to which the corresponding rule of G_1 may be applied. Otherwise, if the substring u and an occurrence of w overlap, then the overlap must be at the left or right end of u since $R_G > B_G$. But for some j this occurrence of w in y corresponds to a substring $w(\$w)^j$ of z. Thus, in G_1, the corresponding rule may be applied to z at the leftmost or right most w of $w(\$w)^j$, as appropriate. With these observations, the proof is straightforward.

<u>Claim 2</u>. $M(L(G_1) \cap S) \subseteq L(G)$.

Not every derivation in G_1 of strings in $L(G_1) \cap S$ can be imitated directly in G, since there are strings z such that $\gamma \to \beta w\$$ is applicable to z while $\gamma w \to \beta w$ is not applicable to $M(z)$. However, by rearranging derivations of G_1, we will show that every $y \in L(G_1) \cap S$ has some derivation which can be imitated in G. That is, for any $y \in S \cap L(G_1)$, there is a derivation of y in which the rule $\gamma \to \beta w\$$ is applied only when context w has already been generated to the right of γ. Such a derivation is "almost" a derivation of G, and it is easy to construct in G a derivation of $M(y)$ which imitates it.

First we show that if $y \in L(G_1) \cap S$, then there is a derivation $X \Rightarrow \theta_1 \Rightarrow \theta_2 \Rightarrow \ldots \Rightarrow \theta_n = y$ in G_1 such that each $\theta_i \in S$.

Labeling the $p \geq 0$ occurrences of $\$$ in x, we may rewrite y as $y = y_1\$_1 y_2 \$_2 \ldots y_p \$_p y_{p+1}$ where each y_i is in V^* and each $\$_i$ is an occurrence of $\$$. Suppose that for some j, $1 \leq j \leq p$, $X = \Gamma_0 \Rightarrow \Gamma_1 \Rightarrow \ldots \Rightarrow \Gamma_m = y$ is a derivation in G_1 such that in some step k, $\$_j$ is generated without context w to its right. In particular, $\Gamma_{k-1} = r\gamma t \underset{G_1}{\Rightarrow} r\beta w\$_j t = \Gamma_k$ with $t \notin wV_1^*$ and $r,t \in V_1^*$.

Now, since $y \in S$ we have $\Gamma_k = r\beta w\$_j t \underset{G_1}{\Rightarrow} \Gamma_{k+1} \Rightarrow \ldots \Rightarrow v\$_j wu = \Gamma_m = y$ for some v,u in V_1^*. Since $\$$ does not appear in the left side of any rule in R_1, v and wu must be generated independently so that $r\beta w \underset{G_1}{\overset{*}{\Rightarrow}} v$ and $t \underset{G_1}{\overset{*}{\Rightarrow}} wu$. Thus,

$$X = \Gamma_0 \Rightarrow \ldots \Rightarrow \Gamma_{k-1} = r\gamma t \overset{*}{\Rightarrow} r\gamma wu \Rightarrow r\beta w\$_j wu \overset{*}{\Rightarrow} v\$_j wu = y$$

is a derivation of y in which $\$_j$ never occurs without context w immediately to the right. For $i < j$, if $\$_i$ always appears with context w in the original derivation, it will still always appear with context w in the new derivation. Thus, the above argument may be applied inductively to the least j such that $\$_j$ is generated without context w to the right. The eventual result is a derivation of y in which no $\$_i$ appears without context w, for $1 \leq i \leq p$.

Next we claim that any derivation in G_1 which at each step produces a word in S can be imitated by a derivation in G. That is, if $X = \theta_0 \Longrightarrow \theta_1 \Rightarrow \ldots \Rightarrow \theta_n$ for $n \geq 0$ is a derivation in G_1 with each θ_i in S, then $X = M(\theta_0) \Longrightarrow M(\theta_1) \Longrightarrow \ldots \Longrightarrow M(\theta_n)$ is a derivation in G.

Note that for any string $usv \in S$ with $u,v \in V_1^*$ and $s \in V^*$, $M(usv)$ contains a substring s (at a point corresponding to the occurrence of s in usv). Thus, for any rule $s \to t$ of G_1, if $s \to t$ is applicable to θ_i it can also be applied at a corresponding point in $M(\theta_i)$. Also, $\gamma \to \beta w\$$ is applicable to θ_i only if w occurs to the right of γ, so that $\gamma w \to \beta w$ is applicable to $M(\theta_i)$. Thus, for any rule applied to θ_i in G_1, a corresponding rule can be applied to a corresponding place in $M(\theta_i)$. With this observation, the proof of the claim by induction on the length of the derivation in G_1 is easy.

The two parts of the above proof show that for any $x \in S \cap L(G_1)$, there is a derivation of $M(x)$ in G. Therefore, $M(S \cap L(G_1)) \subseteq L(G)$.

Case 2. $L_G \leq B_G$ and $R_G \leq B_G$.

As for Case 1, the proof of Case 2 is based on the fact that if the initial portion of a derivation generates a sentential form with a sufficiently long terminal substring w, then subsequent steps must apply rules independently to the segments to the left and right of w. In this case, "sufficiently long" means a length of at least $B_G + 1$. For a terminal string w of that length, no single rule of G can span an entire string AwB, where A and B are arbitrary symbols.

We will construct G_1 by partitioning one of the rules of G into two rules. The rule we will choose to partition will be one which generates a terminal string of length at least $B_G + 1$.

Let $T = \{\alpha \to \beta \in R \mid \alpha \text{ contains at least 2 nonterminals}\}$. Now $T \neq \emptyset$, since otherwise, by definition of B_G, $L_G = R_G \Rightarrow B_G = 0$ so that $N(G) = 0$. So choose from T a rule

$$\alpha_0 Y_1 \alpha_1 \cdots \alpha_{n-1} Y_n \alpha_n \to \alpha_0 \gamma_1 \alpha_1 \cdots \alpha_{n-1} \gamma_n \alpha_n$$

where each $\alpha_i \in \Sigma^*$, each $Y_i \in V - \Sigma$, each $\gamma_i \in V^*$, and for some j, $1 \leq j \leq n-1$, $|\alpha_j| = B_G$. Since G is terminal bounded, the right side of this rule must contain a terminal substring w of length greater than B_G. But $B_G \geq L_G$ and $B_G \geq R_G$. Consequently, w is longer then any string of terminals appearing in the left side of the rule.

Since w is longer than any of the α_i's, some part of at least one of the γ_i's, say γ_k, must be included in w (γ_k may be the empty string). We assume without loss of generality that $k < n$, since the case $k = n$ is symmetric to the case $k = 1$. Specifically, for some k, $1 \leq k < n$, and some $\sigma, \rho, \beta_1, \beta_2 \in V^*$, we may rewrite the rule as $\sigma Y_k \alpha_k Y_{k+1} \rho \to \beta_1 w \beta_2$ where $|\beta_1| < |\alpha_0 \gamma_1 \cdots \gamma_{k-1} \alpha_{k-1} \gamma_k|$. Let \$ and ¢ be distinct new symbols. Let $\Sigma_1 = \{\$, ¢\}$ and $V_1 = V \cup \Sigma_1$. Define a new set of rules $R' = \{\sigma Y_k \to \beta_1 w \$, \alpha_k Y_{k+1} \rho \to \alpha_k ¢ w \beta_2\}$. Observe that by splitting the original rule at α_k and putting an α_k in the right side of the second rule, we have constructed new rules which satisfy the condition that terminals may not be rewritten in grammar rules. Thus, letting $R_1 = (R \cup R') - \{\sigma Y_k \alpha_k Y_{k+1} \rho \to \beta_1 w \beta_2\}$ we may obtain a grammar $G_1 = (V_1, \Sigma \cup \Sigma_1, R_1, X)$. Clearly, G_1 is terminal bounded. Also, since the two rules in R' partition the left side of our selected single rule, which is omitted from R_1, $N(G_1) < N(G)$.

In G, when the rule $\sigma Y_k \alpha_k Y_{k+1} \rho \to \beta_1 w \beta_2$ is applied, the string w is available for use as context by either β_1 or β_2. Also, no single rule of G contains all of w in its left side, so rules must be applied independently to the left and right of w. Replacing this rule by the two rules of R' ensures that if both rules are applied to a string $\tau \sigma Y_k \alpha_k Y_{k+1} \rho \nu$ then the resulting string $\tau \beta_1 w \$ \alpha_k ¢ w \beta_2 \nu$ has w available for use as context by either β_1 or β_2. Thus, if we can restrict our attention to derivations of G_1 in which the two rules of R' are always applied together, and then erase the extra strings $\$\alpha_k¢w$ generated by rules of R', we will be able to obtain exactly the strings in $L(G)$.

Let S be the regular set $(V \cup \{\$\alpha_k¢w\})^*$. Intersection of S with $L(G_1)$ selects those words of $L(G_1)$ which have derivations in which the two rules of R' always occur together.

The extra strings $\$\alpha_k¢w$ used by S to match occurrences of rules in R_1 will be erased by a gsm M which acts as follows on input from V_1^*. In its initial mode of operation, M outputs each symbol it reads. But when M reads the symbol $\$$, it enters an erasing mode in which it erases a string of the form $\$\alpha_k¢w$, thereafter returning to the initial mode of operation. If $\$$ is not followed by $\alpha_k¢w$, M's action is undefined. (Note that for any string in S, $\$$ will always be followed by $\alpha_k¢w$.)

As for Case 1, we will merely sketch the proof that $M(L(G_1) \cap S) = L(G)$.

Claim 3. $L(G) \subset M(S \cap L(G_1))$.

Sketch of the proof. The proof of this inclusion depends on the fact that the whole string w never occurs in the left side of any rule of R. Thus, after a substring w is generated in G, no single context-rule can span all of w; each rule can be applied only at one end or the other of w. Therefore, to construct a derivation in G_1 to imitate a derivation of G, whenever $\sigma Y_k \alpha_k Y_{k+1} \rho \to \beta_1 w \beta_2$ is applied in G, the two rules of R' are applied in G_1 to generate a substring $\beta_1 w \$ \alpha_k ¢ w \beta_2$. Then, whenever a context rule is applied to the left (right) end of w in G, it is applied to the w left of $\$$ (right of $¢$) in G_1. With this observation, the proof of the inclusion is a straightforward induction on n upon the assertion that if $X \overset{n}{\underset{G}{\Longrightarrow}} y$ then there exists a string $z \in S$ such that $X \overset{*}{\underset{G_1}{\Longrightarrow}} z$ and $M(z) = y$.

Claim 4. $M(S \cap L(G_1)) \subseteq L(G)$.

Sketch of the proof. Not every derivation in G_1 of a string in $L(G_1) \cap S$ can be imitated directly in G, since in G_1 the two rules of R' can occur independently. However, we will show that every $y \in L(G_1) \cap S$ has some derivation which can be imitated by a derivation of $M(y)$ in G. That is, y has a derivation in which each application of the rule $\sigma Y_k \to \beta_1 w \$$ is paired with an adjacent application of the rule $\alpha_k Y_{k+1} \rho \to \alpha_k ¢ w \beta_2$ in the next step.

First we need to label the occurrences of \$ and ¢ in y so that we can rewrite y as

$$y = y_0 \$_1 \alpha_k ¢_1 y_1 \$_2 \alpha_k ¢_2 \ldots y_{p-1} \$_p \alpha_k ¢_p y_p$$

where each y_j is in V^* and each $\$_j$ or $¢_j$ is an occurrence of \$ or ¢, respectively.

Now we claim that if $y \varepsilon L(G_1) \cap S$, then there exists a derivation $X \Rightarrow \theta_1 \Rightarrow \ldots \Rightarrow \theta_n = y$ such that for all j, $¢_j$ first appears in θ_{r+1} if $\$_j$ first appears in θ_r.

For suppose that for some j, $1 \leq j \leq p$, $X \Rightarrow \Gamma_1 \Rightarrow \Gamma_2 \ldots \Rightarrow \Gamma_m = y$ is a derivation of y in G_1 such that $¢_j$ is not generated in the step following $\$_j$. We will assume without loss of generality that $¢_j$ is generated at least two steps after $\$_j$, since the proof when $¢_j$ is generated before $\$_j$ is similar. Thus the derivation in G_1 is

$$X \stackrel{*}{\Rightarrow} r\sigma Y_k s \Rightarrow r\beta_1 w \$_j s \stackrel{*}{\Rightarrow} t\$_j \alpha_k Y_{k+1} \rho u \Rightarrow t\$_j \alpha_k ¢_j w \beta_2 u \stackrel{*}{\Rightarrow} v\$_j \alpha_k ¢_j z = y$$

for some $r,s,t,u,v,z \varepsilon V_1^*$.

Since \$ and ¢ do not appear in the left side of any rule of G_1, we know that $s \stackrel{*}{\Rightarrow} \alpha_k Y_{k+1} \rho u$, $r\beta_1 w \stackrel{*}{\Rightarrow} t \stackrel{*}{\Rightarrow} v$ and $w\beta_2 u \stackrel{*}{\Rightarrow} z$. Thus we can obtain in G_1 a new derivation

$$X \stackrel{*}{\Rightarrow} r\sigma Y_k s \stackrel{*}{\Rightarrow} r\sigma Y_k \alpha_k Y_{k+1} \rho u \Rightarrow r\beta_1 w \$_j \alpha_k Y_{k+1} \rho u \Rightarrow r\beta_1 w \$_j \alpha_k ¢_j w \beta_2 u$$
$$\stackrel{*}{\Rightarrow} v\$_j \alpha_k ¢_j z = y.$$

Thus, in the new derivation $¢_j$ first appears in the step following the one which generates $\$_j$. Also, for $i < j$, if $¢_i$ is generated immediately after $\$_i$ in the original derivation, it will still be generated immediately after $\$_i$ in the new derivation. Therefore, the above argument may be applied inductively to the least j such that $¢_j$ is not generated immediately after $\$_j$. The eventual result will be a derivation in which $¢_i$ first appears in the step after the one in which $\$_i$ is generated, for $1 \leq i \leq p$.

Next we need to show that if in a derivation of y in G_1, $\$_i$ is always generated immediately before $¢_i$ is generated, for $1 \leq i \leq p$, then there is a derivation of $M(y)$ in G which "imitates" the derivation of y in G_1. In particular, if $X = \theta_0 \overset{*}{\Rightarrow} \theta_1 \overset{*}{\Rightarrow} \theta_2 \overset{*}{\Rightarrow} \ldots \overset{*}{\Rightarrow} \theta_m = y$ is such a derivation in G_1, where for each $i < m$, $\theta_i \overset{*}{\Rightarrow} \theta_{i+1}$ either via a single rule of R or via consecutive applications of the two rules in R', then $X = M(\theta_0) \Rightarrow M(\theta_1) \Rightarrow \ldots \Rightarrow M(\theta_m) = M(y)$ is a derivation in G.

The above assertion is proved by an induction on m which is based on the following argument. If uyv is a string in V_1^* with $y \in V^*$, then $M(uyv)$ has a substring y corresponding to the occurrence of y in uyv. Thus, if $\theta_i \Rightarrow \theta_{i+1}$ via a rule $y \to z$ of R_1, then $y \to z$ may be applied to $M(\theta_i)$ to produce $M(\theta_{i+1})$. If $\theta_i \overset{2}{\Rightarrow} \theta_{i+1}$ via applications of the two rules of R', then θ_i contains a substring $\sigma Y_k \alpha_k Y_{k+1} \rho$ and $M(\theta_i) \Rightarrow M(\theta_{i+1})$ via the rule $\sigma Y_k \alpha_k Y_{k+1} \rho \to \beta_1 w \beta_2$. Thus $M(\theta_i) \Rightarrow M(\theta_{i+1})$ in G.

From the above, we have shown that for every $y \in L(G_1) \cap S$, $X \overset{*}{\Rightarrow} M(y)$ in G, and thus $M(L(G_1) \cap S) \subseteq L(G)$. $\square$

Section 3.

In this section we show that six previous results are corollaries of Theorem 1. The first four corollaries are simply sub-cases of the theorem for terminal-bounded grammars.

Corollary 1. (Ginsburg and Greibach, 1966) If G is a grammar such that each rule is of the form $\alpha \to \beta$ with $\alpha \in (V - \Sigma)^+$ and $\beta \in V^*\Sigma V^*$, then $L(G)$ is context-free.

Corollary 2. (Book, 1972) If G is a type 0 grammar such that each rule is of the form $\alpha \to \beta$ with $\alpha \in \Sigma^*(V - \Sigma)\Sigma^*$ then $L(G)$ is context-free.

Corollary 3. (Book, 1972) If G is a type 0 grammar such that each rule is of the form $\alpha Z\beta \to \alpha\gamma\beta$ with $|\alpha| \geq |\beta|$ and $\alpha \in \Sigma^*$, then $L(G)$ is context-free.

The following corollary is presented as a conjecture by Book (1972).

<u>Corollary 4</u>. If G is a type 0 grammar such that each rule is of the form $\alpha Z\beta \rightarrow \alpha\gamma\beta$ where either

(i) $\alpha \in \Sigma^*$ and $|\alpha| \geq |\beta|$ or

(ii) $\beta \in \Sigma^*$ and $|\beta| \geq |\alpha|$

then $L(G)$ is context-free.

<u>Corollary 5</u>. (Hibbard, 1966) Let V be a set of symbols on which a partial ordering $<$ is defined. Let $G = (V,\Sigma,R,X)$ be a grammar in which each rule is of the form $A_1A_2\ldots A_n \rightarrow B_1B_2\ldots B_m$ where each $A_i \in V - \Sigma$, and for some k, $1 \leq k \leq m$, and all i, $1 \leq i \leq n$, $A_i < B_k$. Then $L(G)$ is context-free.

<u>Proof</u>. We may assume without loss of generality that $<$ is a total ordering on V, and that X is the least element in this total ordering. Also, we may assume that for every $\alpha \rightarrow \beta \in R$, either $\beta \in (V - \Sigma)^*$ or $\beta \in \Sigma$. We will prove the corollary by encoding the total ordering into strings of terminals. The encoding is complicated by the fact that terminals may not be rewritten in grammar rules.

Let d, E, and X_1 be new symbols, and let $\Sigma_1 = \Sigma \cup \{d\}$, $V_1 = V \cup \{d,X_1,E\}$. Define p to be the maximum length of the left side of any rule in R. Let h be the homomorphism from V^* to V_1^* determined by $h(A) = AdP^i$ for $A \in V$ if A is greater than exactly $i \geq 0$ elements. Construct a new set of rules

$$R_1 = \{X_1 \rightarrow XdE,\ E \rightarrow d\} \cup \{h(\alpha)Y \rightarrow h(\beta)Y \mid \alpha \rightarrow \beta \in R \text{ and } Y \in V_1-\Sigma_1\}.$$

Now, if $\alpha \to \beta \in R$ and the maximum symbol of α is greater than exactly i symbols, then the maximum symbol of β is greater than at least $i+1$ symbols. Therefore, for any $Y \in V_1 - \Sigma_1$, the left side of $h(\alpha)Y \to h(\beta)Y$ contains at most $p(p^i)$ d's, while the right side contains a string of at least p^{i+1} d's. Suppose that for some $n,m > 0$, some $i_1,\ldots,i_n,j_1,\ldots,j_m$ and some $A_1,\ldots,A_n,B_1,\ldots,B_m \in V_1$, $h(\alpha)Y = A_1 d^{i_1} A_2 d^{i_2} \ldots A_n d^{i_n} Y$ and $h(\beta)Y = B_1 d^{j_1} B_2 d^{j_2} \ldots B_m d^{j_m} Y$. If the longest string of d's in $h(\beta)Y$ is d^{j_k}, then the rule may be interpreted as follows. All of the symbols to the left of d^{j_k} and possibly some of the d's in d^{j_k} are generated by A_1, while all of the symbols to the right of d^{j_k} are generated by Y. Each d occurring in the left side of the rule is one of the d's in d^{j_k}, while each nonterminal other than A_1 and Y is rewritten as the empty string. Thus, the new rule satisfies the condition that terminals may not be rewritten in grammar rules.

Let $G_1 = (V_1,\Sigma_1,R_1,X_1)$ be a new grammar. Obviously G_1 is terminal-bounded, so by the theorem $L(G_1)$ is context-free. Let h_1 be the homomorphism from V_1^* to V^* which erases d and is the identity on other symbols. Notice that if $\alpha \Rightarrow \beta$ in G then $h(\alpha)E \Rightarrow h(\beta)E$ in G_1. Also, if $\alpha E \Rightarrow \beta E$ in G_1, then $h_1(\alpha) \Rightarrow h_1(\beta)$ in G_1. Since in G_1 an E can only be rewritten as a d, it is clear that $h_1(L(G_1)) = L(G)$. Thus $L(G)$ is context-free. $\square$

Definition 4. Let $G = (V,\Sigma,R,X)$ be a grammar. If $\alpha \in \Sigma^*$, $\sigma \to \gamma \in R$, and $\beta \in V^*$, write $\alpha\sigma\beta \overset{L}{\Rightarrow} \alpha\gamma\beta$ and $\beta\sigma\alpha \overset{R}{\Rightarrow} \beta\gamma\alpha$. Define $\mathrm{Left}(G) = \{w \in \Sigma^* \mid X \overset{L}{\Rightarrow} \theta_1 \overset{L}{\Rightarrow} \theta_2 \Rightarrow \ldots \overset{L}{\Rightarrow} \theta_n = w\}$ and $\text{Two-way}(G) = \{w \in \Sigma^* \mid X \Rightarrow \theta_1 \Rightarrow \theta_2 \Rightarrow \ldots \Rightarrow \theta_n = w$ and for $i = 2,\ldots,n$, either $\theta_{i-1} \overset{L}{\Rightarrow} \theta_i$ or $\theta_{i-1} \overset{R}{\Rightarrow} \theta_i\}$.

Corollary 6. (Matthews, 1967) If $G = (V,\Sigma,P,X)$ is a grammar, then Two-way(G) is context-free.

Proof. The following construction obtains from G a terminal-bounded grammar G_1 with $L(G_1) = \text{Two-way}(G_1)$, such that for some homomorphism h, $h(L(G_1)) = \text{Two-way}(G)$. In order to force $L(G_1) = \text{Two-way}(G_1)$, markers are added to mark the leftmost and rightmost nonterminals at any point in the derivation. In order to make G_1 terminal-bounded, extra terminal context is added on the left or right side of each rule; the terminal context will not interfere with application of rules since each rule can be applied only when there are no nonterminals on at least one side.

Define q to be the length of the longest left side of a rule of P. Let d, L, R, and X_1 be new symbols and let $\Sigma_1 = \Sigma \cup \{d\}$, $V_1 = V \cup \{d,L,R,X_1\}$. Construct new sets of rules

$$P_1 = \{X_1 \to d^q LXRd^q, \quad L \to e, \quad R \to e\}$$

$$P_2 = \{yuL\alpha v \to ywL\beta x, \quad u\alpha Rvy \to w\beta Rxy, \quad yuL\alpha Rv \to ywL\beta Rx \mid u\alpha v \to w\beta x \in P \text{ and } u,v,w,x,y \in \Sigma_1^* \text{ and } \alpha,\beta \in \{e\} \cup (V-\Sigma)(\Sigma^*(V-\Sigma))^* \text{ and } |y| = q\}.$$

Define a new grammar $G_1 = (V_1, \Sigma_1, P_1 \cup P_2, X_1)$.

Note that L can appear in a sentential form of G_1 only when no nonterminals appear to its left, while R appears only when no nonterminals are to the right. Since every rule of P_1 other than the single rule with left side X_1 contains either an L or an R, $L(G_1) \subseteq \text{Two-way}(G_1) \subseteq L(G_1)$. Thus $L(G_1) = \text{Two-way}(G_1)$. Also, since q is the length of the longest left side of any rule of P, and the left side of each rule of G_1 has a terminal string of length at least q at one end, it is clear that the left side of each rule contains a terminal string longer than any terminal string which appears between two nonterminals. Thus G_1 is terminal-bounded (in fact, G_1 satisfies the restriction given in Corollary 4).

Let h be a homomorphism which erases d but is the identity on other symbols. We claim that $h(L(G_1)) = L(G)$. For suppose P contains a rule $u\alpha v \to w\beta x$ where $u,v,w,x \in \Sigma^*$ and α and β both begin and end with nonterminals. During a derivation of G, if an application of this rule to a string z rewrites the leftmost nonterminal of z (but not the rightmost nonterminal of z) then $yuL\alpha v \to ywL\beta x$ can be applied to a corresponding string in G_1, where y is the terminal string of length q to the left of u. Similarly, if in G $u\alpha v \to w\beta x$ rewrites the rightmost nonterminal of z, then for the appropriate $y \in \Sigma^*$, $u\alpha Rvy \to w\beta Rxy$ can be applied in G_1 at the rightmost nonterminal. If $u\alpha v \to w\beta x$

rewrites both the leftmost and rightmost nonterminals, then $yuL\alpha Rv \to ywL\beta Rx$ (for the appropriate y) can be applied in G_1. Thus it is clear that $\text{Two-way}(G) \subseteq h(L(G_1))$. On the other hand, every rule of G_1 (except those in P_1) imitates a rule in G; thus $h(\text{Two-way}(G_1)) = h(L(G_1)) \subseteq \text{Two-way}(G)$. Therefore $L(G_1)) = \text{Two-way}(G)$. Now since G_1 is terminal-bounded, $L(G_1)$ is context-free. Consequently $\text{Two-way}(G)$ is context-free. □

Corollary 7. (Matthews, 1964) If $G = (V,\Sigma,R,X)$ is a grammar, then $\text{Left}(G)$ is context-free.

Proof. The proof of Corollary 7 is similar to that of Corollary 6.

The various corollaries are closely related to each other. In particular, Corollaries 2 and 3 (Book, 1972) are subcases of Corollary 4. The Ginsburg-Greibach (1966) result, Corollary 1, can be obtained easily from Corollary 5 by defining a partial ordering on symbols in which every terminal is greater than every nonterminal. Ginsburg and Greibach (1966) show that Corollary 6 follows from Corollary 1; the above proof of Corollary 6 shows that it also follows from Corollary 4. Similarly, Corollary 7 can be derived from both Corollaries 1 and 3. Since Book (1972) obtains Corollary 3 from Corollary 7, we see that Corollaries 3 and 7 are equivalent. On the other hand, it does not seem easy to derive Corollary 4 or Corollary 5 from the other corollaries; they are the strongest of the corollaries.

All of the corollaries are easily derived from the theorem presented in this paper. Consequently, the proof of this theorem provides insight into the basic mechanism underlying all of these results. The two constructions in this paper depend on the fact that when a sufficiently long terminal substring is generated during a derivation, then subsequent rules must be applied independently to the segments on either side of the terminal substring; the rules cannot pass information from one side of the segment to the other. In Case 1, this blockade effect of terminal strings means that a rule with long terminal context can be replaced by a rule with shorter context. Application of the new rule requires a "guess" that the additional context is present; the guess is checked by means of new symbols which mark applications of the new rule. In Case 2, the blockade effect makes it possible to replace a rule which generates a long terminal string by two rules which partition the original rule. Application of the new rule involves a "guess" that the other is also applied, and new symbols which mark applications of the new rules are used to check the guesses.

Thus, the theorem presented here and all of its corollaries follow from the ability of terminal strings to block the use of context to pass information between segments of a string. Therefore, only one basic mechanism forcing a grammar to generate a context-free language has been isolated in the new theorem and all the previous results.

References

Book, R. (1972), Terminal context in context-sensitive grammars, SIAM Journal of Computing 1, 20-30.

Ginsburg, S. (1966), The Mathematical Theory of Context-free Languages, McGraw Hill, New York

Ginsburg, S. and Greibach, S. (1966), Mappings which preserve context-sensitive languages, Information and Control 9, 563-582.

Hibbard, T. (1966), Scan Limited Automata and Context Limited Grammars, Doctoral dissertation, University of California at Los Angeles.

Matthews, G. (1963), Discontinuity and asymmetry in phrase structure grammars, Information and Control 6, 137-146.

Matthews, G. (1964), A note on asymmetry in phrase structure grammars, Information and Control 7, 360-365.

Matthews, G. (1967), Two-way languages, Information and Control 10, 111-119.