TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 279, Number 1, September 1983

GROUPS AND SIMPLE LANGUAGES

BY

ROBERT H. HARING-SMITH[1]

ABSTRACT. With any finitely generated group presentation, one can associate a formal language (called the reduced word problem) consisting of those words on the generators and their inverses which are equal to the identity but which have no proper prefix equal to the identity. We show that the reduced word problem is a simple language if and only if each vertex of the presentation's Cayley diagram has only a finite number of simple closed paths passing through it. Furthermore, if the reduced word problem is simple, then the group is a free product of a free group of finite rank and a finite number of finite groups.

Let $\pi = \langle X; R \rangle$ be a finitely generated (f.g.) presentation of a group G, and let $\Sigma = X \cup X^{-1}$ be the set of generators and their inverses. Define the *word problem* of π, denoted by $\mathrm{WP}(\pi)$, to be the set of all words on Σ which are equal to the identity element of G. Let the *reduced word problem* of π, denoted by $\mathrm{WP}_0(\pi)$, be the subset of $\mathrm{WP}(\pi)$ consisting of those words having no proper prefix equal to the identity. The general problem arises of determining the relationship between the properties of the group G and those of the formal languages $\mathrm{WP}(\pi)$ and $\mathrm{WP}_0(\pi)$.

A. V. Anisimov [**1**] was the first to attempt a classification of f.g. groups according to the position of their word problems in the Chomsky hierarchy of languages. He found that if one f.g. presentation of a group has a regular (resp., context-free) word problem, then all f.g. presentations of that group have regular (resp., context-free) word problems. Furthermore, an f.g. group has a regular word problem (resp., reduced word problem) if and only if it is finite. Anisimov proved some closure properties for the class of groups with context-free word problems, but was unable to characterize this class in algebraic terms.

Recent work by D. E. Muller and P. E. Schupp [**3**] shows that an f.g. group is an extension of a free group by a finite group if and only if it has a context-free word problem and is accessible (that is, satisfies a certain finite chain condition). If, as C. T. C. Wall has conjectured, all f.g. groups are accessible, then Muller and Schupp have characterized groups with context-free word problems. Their work also implies that any group having a context-free word problem, in fact, has a deterministic context-free word problem.

Received by the editors March 11, 1982 and, in revised form, October 6, 1982. Parts of this paper were presented to the AMS at the regional meeting in Norman, Oklahoma, on March 18, 1983.

1980 *Mathematics Subject Classification*. Primary 20E06, 20F10, 68F05.

Key words and phrases. Word problem for groups, free product of groups, context-free language, simple language, Cayley diagram, partial group.

[1]The results in this paper are taken from the author's Ph.D. thesis, written under the supervision of Professor Paul Schupp while the author was a University Fellow of the University of Illinois at Urbana-Champaign.

0002-9947/82/0000-1024/$07.00

This paper considers simple languages, which form a proper subclass of the deterministic context-free languages. Every simple language is prefix-free, so the reduced word problem is a more appropriate object for study in this regard than is the word problem. Although one presentation of a group may have a simple reduced word problem and another not, there are interesting geometric and algebraic characterizations of groups having some presentation with simple reduced word problem. The main results of this paper are

THEOREM 1. *An f.g. presentation π has a simple reduced word problem if and only if through each vertex in the Cayley diagram of π there are only a finite number of simple closed paths.*

THEOREM 2. *A group G has an f.g. presentation with a simple reduced word problem if and only if G is the free product of a free group of finite rank and a finite number of finite groups.*

Work by J. R. Stallings [**4**] plays an important role in the proof of Theorem 2, which proceeds in several stages. First, with any f.g. presentation π of a group G, one associates a partial group H. Second, if $\mathrm{WP}_0(\pi)$ is simple, then H is shown to be finite. Finally, Stallings' characterization of free products is invoked to determine the structure of G. The converse if proven by constructing a simple grammar for the reduced word problem of a particular presentation of G.

I. Preliminary definitions. This section presents basic terminology and facts concerning languages and presentations.

Let Σ be a finite set of symbols or *letters*—the *alphabet* for a language. A *word* on Σ is an expression of the form $x_1x_2 \cdots x_n$, where n is a nonnegative integer and each x_i is an element of Σ. The word containing no symbols is the *empty word* and is denoted by Λ (with the stipulation that "Λ" not be a member of Σ). Let Σ^* be the set of all words on Σ. A *language* on Σ is any subset of Σ^*.

If w and v are words on Σ—say, w is $x_1x_2 \cdots x_n$ and v is $y_1y_2 \cdots y_m$— then the expression $w \equiv v$ indicates that w and v are identical strings. This means $m = n$ and, for each i, the letters x_i and y_i are the same symbol. The *concatenation* of w and v is the word $wv \equiv x_1x_2 \cdots x_ny_1y_2 \cdots y_m$. A *subword* of w is a word $u \equiv x_ix_{i+1} \cdots x_j$, where $1 \leqslant i \leqslant j \leqslant n$. The empty word is regarded as a subword of every word. The subword u is a *prefix* of w if $i = 1$ and a *suffix* of w if $j = n$. A subword, prefix, or suffix of w is *proper* if it is neither empty nor identical to w. In later sections, a word u may be characterized as a subword of some element of a set W of words. It is convenient to say simply that u is a *subword of* W. Similarly, a word may be described as a prefix of W or a suffix of W. The *length* of the word w—written $|w|$ —is the number of letters in w.

A *group presentation* is a pair $\pi = \langle X; R \rangle$ where X is a set of letters called *generators* and R, the set of *defining relators*, consists of words on the generators and their inverses. Only finitely generated presentations (those in which X is finite) will be considered. The presentation π determines a group in the following way: Let F be the free group with basis X and let N be the smallest normal subgroup of F which contains R (that is, N is the normal subgroup of F generated by R). Then π presents the group $G = F/N$. Since certain properties dealt with in later sections may be possessed by some but not all presentations of a given group, the distinction between group and group presentation will be carefully maintained.

The notion of a group presentation is linked to that of a formal language as follows: Let X^{-1} be a set of symbols $\{x^{-1}: x \in X\}$ disjoint from X. The set $\Sigma = X \cup X^{-1}$ serves as the alphabet for two languages associated with the presentation π.

DEFINITION. The *word problem* of π, denoted by $\mathrm{WP}(\pi)$, is the set of all words on Σ which are equal to the identity element in the group G presented by π. The *reduced word problem*, $\mathrm{WP}_0(\pi)$, is the set of all nonempty words in $\mathrm{WP}(\pi)$ which have no proper prefix in $\mathrm{WP}(\pi)$. A word on Σ can be regarded as either a word or a group element, and the concatenation of two words can be regarded as either a new string of symbols or the product of two group elements. While $w \equiv v$ indicates that two words are identical, the expression $w = v$ implies only that they represent the same element of the group.

Many languages can be described by means of grammars, which are formal systems for generating the words in the language. To be precise, a *grammar* is a four-tuple $\Gamma = (V, \Sigma, P, S)$ where V is a nonempty set called the *vocabulary*, Σ is a nonempty subset of V called the *terminal alphabet*, S is an element of $V - \Sigma$ called the *start symbol*, and P is a finite set of *productions*. Each production has the form $\alpha \to \beta$ in which α and β are elements of V^* and α contains at least one symbol from $V - \Sigma$. The shorthand notation $\alpha \to \beta_1 \mid \beta_2 \mid \cdots \mid \beta_n$ can be used to represent several productions $\alpha \to \beta_1, \ldots, \alpha \to \beta_n$ having the same word α on the left. Let $N = V - \Sigma$ denote the set of *nonterminal* symbols, or *variables*.

A grammar generates words in the following manner: Let α' and β' be words on V. If there are words $\alpha_1, \alpha_2, \alpha$, and β in V^* such that $\alpha' \equiv \alpha_1\alpha\alpha_2$, $\beta' \equiv \beta_1\beta\beta_2$, and $\alpha \to \beta$ is in P, then α' *directly generates* β' (or β' can be *derived directly* from α'). More generally, the word α' *generates* β' (or β' can be *derived* from α') if there is a sequence $\alpha_0, \alpha_1, \ldots, \alpha_n$ of words on V such that $\alpha' \equiv \alpha_0$, $\beta' \equiv \alpha_n$, and α_i directly generates α_{i+1} for $i = 0, 1, \ldots, n-1$. That α' generates β' is denoted by $\alpha' \overset{*}{\Rightarrow} \beta'$. The sequence $\alpha_0, \alpha_1, \ldots, \alpha_n$ is a *derivation* of β' from α'.

DEFINITION. The language $L(\Gamma)$ generated by a grammar $\Gamma = (V, \Sigma, P, S)$ is the set of all words on the terminal alphabet Σ which can be derived from the start symbol S. That is,

$$L(\Gamma) = \{w \in \Sigma^*: S \overset{*}{\Rightarrow} w\}.$$

A grammar $\Gamma = (V, \Sigma, P, S)$ is *context-free* if each production has the form $A \to \alpha$, where A is a variable and α is a word on V. Notice that the form a production may take has been restricted so that the replacement of a variable in a string of symbols does not depend on its location, or context, in the string. A language L is *context-free* if there is a context-free grammar which generates L.

As is shown in any text on formal languages (e.g., Harrison **[2]**), even more restricted types of grammars are capable of generating all context-free languages. The use of Greibach normal form clarifies the relationship between simple languages and context-free languages. A grammar $\Gamma = (V, \Sigma, P, S)$ is *reduced* if $P = \varnothing$ or if, for every $A \in V$, there are words α and β in V^* and w in Σ^* such that $S \overset{*}{\Rightarrow} \alpha A \beta \overset{*}{\Rightarrow} w$. That is, the grammar contains no variables which do not participate in the derivation of some terminal string. A grammar is in *Greibach normal form* if it is reduced and if each production has one of the forms

(i) $A \to aB_1B_2 \cdots B_n$,

(ii) $A \to a$, or

(iii) $S \to \Lambda$,

where A is a variable, each B_i is a variable other than S, and a is a terminal.

LEMMA 1. *Every context-free language may be generated by a grammar in Greibach normal form.*

In simple grammars, additional restrictions are placed on the sorts of productions which may originate from a particular variable.

DEFINITION. A grammar $\Gamma = (V, \Sigma, P, S)$ in Greibach normal form is *simple* if the following two conditions are satisfied:

(i) For any variable A, terminal a, and words α and β on V, if $A \to a\alpha$ and $A \to a\beta$ are productions, then $\alpha \equiv \beta$.

(ii) If $S \to \Lambda$ is a production, then it is the only production.

A language is *simple* if it can be generated by a simple grammar.

Simple languages have been studied for some time. They can be characterized as those formal languages accepted by a one-state deterministic pushdown automaton which halts when its stack empties. An unusual property of the class of simple grammars is that its equivalence problem is solvable. That is, there is an algorithm which, given two simple grammars, determines if they generate the same language. See Harrison [**2**] for details and further references.

The group theoretic construction most often used in this paper is the free product. A group G is the *free product* of its subgroups $H_1, H_2, \ldots, H_n$ if, for any group K and set of homomorphisms $\{\phi_1, \ldots, \phi_n\}$ where $\phi_i\colon H_i \to K$, there is a unique homomorphism $\phi\colon G \to K$ such that each ϕ_i is the restriction of ϕ to H_i. The expression

$$G = H_1 * H_2 * \cdots * H_n$$

indicates that G is the free product of $H_1, \ldots, H_n$. Each subgroup H_i is a *free factor* of G.

A *normal form* is a sequence $(g_1, \ldots, g_k)$ of nontrivial elements of the free product G such that each g_i is in one of the free factors and consecutive elements g_i, g_{i+1} are in different factors. A normal form $(g_1, \ldots, g_k)$ *represents* the group element $g = g_1g_2 \cdots g_k$. The empty normal form represents the identity element. The basic theorem about elements of a free product is

THE NORMAL FORM THEOREM FOR FREE PRODUCTS. *Each element of a free product can be represented in exactly one way as the product $g_1g_2 \cdots g_k$ of elements in a normal form $(g_1, g_2, \ldots, g_k)$.*

II. Simple grammars for word problems. This section presents several lemmas describing properties of context-free grammars which generate a word problem or reduced word problem. Extensive use is made of a particular type of derivation, the *leftmost derivation*, in which, at each step, it is the leftmost variable to which a production is applied. To be precise, if $\alpha_1 A \alpha_2 \Rightarrow \alpha_1 \beta \alpha_2$ is a step in a leftmost derivation made by using the production $A \to \beta$, then α_1 consists entirely of terminal symbols. Every word in the language generated by a context-free grammar has a leftmost derivation. One can be obtained, for example, simply by changing the order in which productions are used in an arbitrary derivation.

Leftmost derivations from the start symbol in a grammar in Greibach normal form are especially easy to describe. There may be one anomalous leftmost derivation: $S \overset{*}{\Rightarrow} \Lambda$. In the rest, each production introduces precisely one terminal, and no individual production creates a word in which a variable is to the left of a terminal. The outcome of a leftmost derivation of length n must therefore be a word $u\alpha$ in which u is a terminal string of length n and α is a nonterminal string.

If the grammar is simple, the leftmost derivation of $u\alpha$ is unique; at the ith step there is only one way to obtain from the leftmost variable a word beginning with the ith symbol of u. Furthermore, if $u\alpha$ and $uv\beta$ are two words which have leftmost derivations originating with S and a common prefix u, then the first $|u|$ steps in their leftmost derivations are identical—completely determined by u. These remarks constitute most of the proof of the next lemma.

DEFINITION. A language L is *prefix-free* if u and uv are both words in L only if v is empty.

LEMMA 2. *A simple language is prefix-free.*

PROOF. Let L be a language generated by a simple grammar $\Gamma = (V, \Sigma, P, S)$. If u and uv can be derived from S, then, as observed above, their leftmost derivations must be identical for the first $|u|$ steps. Since u is in L, no variables will remain after those steps, so the derivation cannot be extended. Consequently, v must be empty.

The conclusion of the preceding paragraph is valid if u is empty, though not for the same reason. If L contains the empty word, then it consists solely of the empty word and is certainly prefix-free. □

Lemma 2 implies that the word problem is a simple language only in the rather uninteresting case of the empty presentation of the trivial group. For any other presentation, the word problem contains nonempty words, the concatenation of any two of which is an element of the word problem having a proper prefix also in the word problem. Since the reduced word problem is always prefix-free, it is more appropriate for study in connection with simple languages.

The next few lemmas apply to less restricted forms of context-free grammars than simple grammars, but for the last lemmas of this section, the assumption that the grammar is simple is crucial.

LEMMA 3. *Let π be a finitely generated presentation of a group G, and let $\Gamma = (V, \Sigma, P, S)$ be a context-free grammar generating a language L contained in* WP(π). *If α is any variable string arising in the derivation of an element of L, and if y and z are terminal strings that can be derived from α, then $y = z$ in G.*

PROOF. The reason is essentially that y and z can be substituted for each other in elements of the word problem. Let

$$S \stackrel{*}{\Rightarrow} \gamma\alpha\beta \stackrel{*}{\Rightarrow} uvw,$$

where α and β are in V^*, $\gamma \stackrel{*}{\Rightarrow} u$, $\alpha \stackrel{*}{\Rightarrow} v$, and $\beta \stackrel{*}{\Rightarrow} w$. Using the derivations $\alpha \stackrel{*}{\Rightarrow} y$, and $\alpha \stackrel{*}{\Rightarrow} z$, construct two new terminal derivations

$$S \stackrel{*}{\Rightarrow} \gamma\alpha\beta \stackrel{*}{\Rightarrow} uyw \quad \text{and} \quad S \stackrel{*}{\Rightarrow} \gamma\alpha\beta \stackrel{*}{\Rightarrow} uzw.$$

Then $uyw = 1 = uzw$, so $z = y$. □

In a reduced grammar, each variable appears in the derivation of a terminal string, so Lemma 3 implies that the following notation is well defined.

DEFINITION. Let π be a finitely generated presentation of a group G, and let Γ be a reduced context-free grammar generating a subset of $\mathrm{WP}(\pi)$. For any variable A, define $\sigma(A)$ to be the element of G represented by any word derivable from A. By extension, if $\alpha \equiv A_1A_2\cdots A_k$ is a string of variables, let $\sigma(\alpha) = \sigma(A_1)\cdots\sigma(A_k)$. For example, if S is the start symbol of Γ, then $\sigma(S) = 1$.

The next lemma will be used in subsequent proofs to show that the language generated by a given grammar is the reduced word problem of a presentation.

LEMMA 4. *If π is a finitely generated group presentation and L is a prefix-free language such that* $\mathrm{WP}_0(\pi) \subset L \subset \mathrm{WP}(\pi)$, *then* $L = \mathrm{WP}_0(\pi)$.

PROOF. Since $L \subset \mathrm{WP}(\pi)$, any element w in L can be written as the concatenation of words $w_1, w_2, \ldots, w_k$ in $\mathrm{WP}_0(\pi)$. Then both w and its nonempty prefix w_1 are elements of the prefix-free language L, so w and w_1 are identical. It follows that w is in $\mathrm{WP}_0(\pi)$ and, hence, $\mathrm{WP}_0(\pi)$ contains L. The reverse containment is one of the hypotheses, so $L = \mathrm{WP}_0(\pi)$. □

For the remainder of this section, let $\pi = \langle X; R\rangle$ be a finitely generated presentation of a group G. Let $\Gamma = (V, \Sigma, P, S)$ be a simple grammar generating $\mathrm{WP}_0(\pi)$.

LEMMA 5. *Let u and w be nonempty words on Σ, and suppose v is a word in* $\mathrm{WP}(\pi)$. *The word uvw is in* $\mathrm{WP}_0(\pi)$ *if and only if uw is in* $\mathrm{WP}_0(\pi)$ *and no nonempty prefix of uv is in* $\mathrm{WP}(\pi)$.

PROOF. If uvw is in $\mathrm{WP}_0(\pi)$, then no proper prefix of uvw and hence no nonempty prefix of uv can be in $\mathrm{WP}(\pi)$. Furthermore, since $v = 1$, the word uw is in $\mathrm{WP}(\pi)$ and would fail to be in $\mathrm{WP}_0(\pi)$ only if it had a proper prefix equal to the identity. Such a prefix would have one of the following forms:

(i) u_1, where u_1 is a nonempty prefix of u, or

(ii) uw_1, where w_1 is a proper prefix of w.

In either case, the word uvw would have a proper prefix equal to the identity—u_1 in case (i) and uvw_1 in case (ii). Since this is not possible, the word uw is in $\mathrm{WP}_0(\pi)$.

Conversely, if $uw \in \mathrm{WP}_0(\pi)$, then uvw is at least in $\mathrm{WP}(\pi)$. A proper prefix of uvw is either a nonempty prefix of uv, which has no such prefix equal to the identity, or has the form uvw_1, where w_1 is a proper prefix of w. But if $uvw_1 = 1$, then uw_1

would be a proper prefix of uw equal to the identity, contrary to the assumption that $uw \in \mathrm{WP}_0(\pi)$. Therefore, $uvw \in \mathrm{WP}_0(\pi)$. □

LEMMA 6. *For each variable A and terminal x, there is a unique production $A \to x\alpha$, where $\alpha \in N^*$.*

PROOF. Only the existence of the production requires proof; it is unique because Γ is simple. Since Γ is reduced, there is a leftmost derivation

$$S \overset{*}{\Rightarrow} uA\beta \overset{*}{\Rightarrow} uyz, \tag{1}$$

where u, y, and z are terminal strings, β is a variable string, $A \overset{*}{\Rightarrow} y$, and $\beta \overset{*}{\Rightarrow} z$. Consider two cases:

(i) If $ux = 1$, then ux would fail to be in $\mathrm{WP}_0(\pi)$ only if it had a proper prefix in $\mathrm{WP}(\pi)$. Since such a prefix would also be a prefix of uyz and uyz is in $\mathrm{WP}_0(\pi)$, no such prefix exists. Then there is a leftmost derivation $S \overset{*}{\Rightarrow} ux$, which must begin $S \overset{*}{\Rightarrow} uA\beta$ by the discussion preceding Lemma 2. The variable string β is therefore empty, and $A \to x$ is in P.

(ii) If $ux \neq 1$, then Lemma 5 can be used to show that $uxx^{-1}yz \in \mathrm{WP}_0(\pi)$. Let $u \equiv u$, $v \equiv xx^{-1}$, and $w \equiv yz$. The possibility that a nonempty prefix of uv is in $\mathrm{WP}(\pi)$ was handled in case (i), so assume that no such prefix exists. The existence of derivation (1) implies that $uyz \equiv uw$ is in $\mathrm{WP}_0(\pi)$. The word $uvw \equiv uxx^{-1}yz$ is therefore in $\mathrm{WP}_0(\pi)$.

The leftmost derivation of $uxx^{-1}yz$ must begin $S \overset{*}{\Rightarrow} uA\beta$ and can continue only via a production $A \to x\alpha$, so the lemma is proven. □

While in a reduced grammar every variable occurs in some derivation, there are strings of variables which cannot arise in derivations in the (reduced) simple grammar Γ. The next lemma discusses some of those strings.

LEMMA 7. *If α is a nonempty string of variables arising in a leftmost derivation of some terminal string and $\sigma(\alpha) = 1$, then $\alpha \equiv S$.*

PROOF. If α is not S but satisfies the hypotheses of the lemma, then there is a leftmost derivation

$$S \overset{*}{\Rightarrow} t\alpha\beta \overset{*}{\Rightarrow} tuv, \tag{2}$$

where t, u, and v are terminal strings, β is a variable string, $\alpha \overset{*}{\Rightarrow} u$, and $\beta \overset{*}{\Rightarrow} v$. The word t is nonempty since $\alpha \not\equiv S$, the word u is nonempty since the empty word cannot be derived from a nonempty string of variables in Γ, and the word v is nonempty since otherwise tuv would have a proper prefix (i.e., t) equal to the identity.

Assume, without loss of generality, that tuv is the shortest word having a derivation of the form shown in line (2). Since $u = \sigma(\alpha) = 1$ and $tuv \in \mathrm{WP}_0(\pi)$, Lemma 5 implies that $tv \in \mathrm{WP}_0(\pi)$. The leftmost derivation of tv must begin

$S \stackrel{*}{\Rightarrow} t\alpha\beta$, so there are words v_1 and v_2 such that $v \equiv v_1 v_2$, $\alpha \stackrel{*}{\Rightarrow} v_1$, and $\beta \stackrel{*}{\Rightarrow} v_2$. Then the derivation

$$S \stackrel{*}{\Rightarrow} t\alpha\beta \stackrel{*}{\Rightarrow} tv_1v_2$$

has the same form as that displayed on line (2), but tv_1v_2 is shorter than tuv. The lemma follows from this contradiction to the minimality of $|tuv|$. □

The final lemma of this section extends to all variables a property vested in S by virtue of the definition of $\mathrm{WP}_0(\pi)$.

LEMMA 8. *For any variable A and terminal string u, there is a unique leftmost derivation having one of the following forms*:

(i) $A \stackrel{*}{\Rightarrow} u_1$, *where u_1 is a proper prefix of u, or*

(ii) $A \stackrel{*}{\Rightarrow} u\alpha$, *where $\alpha \in N^*$.*

Thus, if $u = \sigma(A)$ but no proper prefix of u is equal to $\sigma(A)$, then $A \stackrel{*}{\Rightarrow} u$.

PROOF. By Lemma 6, there are enough productions in P to continue a derivation of u from A until the end of u is reached or until no variables remain, whichever comes first. If the variables are exhausted first, then a derivation of type (i) occurs. Otherwise, the derivation has form (ii).

If no proper prefix of u is equal to $\sigma(A)$, then a derivation of the first type is not possible. On the other hand, in a type (ii) derivation, if $u = \sigma(A)$ then $\sigma(\alpha) = 1$. By Lemma 7, the word α must be empty, so $A \stackrel{*}{\Rightarrow} u$. □

III. A geometric characterization of presentations with simple reduced word problems. In this section, presentations having simple reduced word problems are characterized by identifying a property of their Cayley diagrams.

A *graph* Δ consists of a set V of *vertices*, a set E of *edges*, functions α: $E \to V$ and ω: $E \to V$ determining the *initial* and *final vertices* of an edge, and a function i: $E \to E$ associating with each edge an *inverse edge*. The function i is required to have the properties that, for all edges e,

(i) $i(i(e)) = e$,

(ii) $\alpha(i(e)) = \omega(e)$, and

(iii) $\omega(i(e)) = \alpha(e)$.

The last two equations say that the initial vertex of $i(e)$ is the final vertex of e, and the final vertex of $i(e)$ is the initial vertex of e. To simplify notation and emphasize the connection which is to be established between graphs and presentations, the inverse edge $i(e)$ shall be denoted by e^{-1}. When depicting graphs using points and curves in a plane, the edges e and e^{-1} shall be represented by the same segment traversed in opposite directions.

The *Cayley diagram* for a presentation $\pi = \langle X; R\rangle$ of a group G is a graph $\Delta(\pi)$ equipped with labelling functions λ: $V \to G$ and μ: $E \to X \cup X^{-1}$ such that:

(i) λ is a bijection,

(ii) for each $v \in V$, the function μ restricted to the set $E(v)$ of edges with initial vertex v is a bijection between $E(v)$ and $X \cup X^{-1}$, and

(iii) for all edges e, the equation $\lambda\alpha(e) \cdot \mu(e) = \lambda\omega(e)$ is true in G.

These conditions imply that the vertices of $\Delta(\pi)$ are in one-to-one correspondence with the elements of G, that for each vertex v and letter a in Σ there is a distinct edge with initial vertex v and label a, and that there is an edge labelled a having initial vertex labelled g_1 and final vertex labelled g_2 if and only if $g_1 a = g_2$ in G.

The following examples illustrate these definitions.

EXAMPLE 1. One presentation of the free product of cyclic groups of orders two and three is $\pi_1 = \langle a, b; a^2, b^3 \rangle$. The Cayley diagrams for the presentations $\langle a; a^2 \rangle$ and $\langle b; b^3 \rangle$ alone are shown below in Diagrams 1 and 2, respectively. A portion of the diagram for the free product appears in Diagram 3. In the full Cayley diagram, the pattern evident in this fragment is repeated indefinitely.

Each edge shown in the diagrams has been labelled by a generator and marked with an arrow to indicate the orientation which corresponds to that label. The inverse edge is represented by the same segment in the plane traversed in the direction opposite the arrow and would be labelled by the inverse generator. The only labelled vertex shown is that corresponding to the identity. Labels on the other vertices are determined by a choice of identity vertex and the labels on the edges. Notice that leaving each vertex in $\Delta(\pi_1)$ are exactly four edges, one for each of the two generators and their inverses.

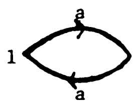

DIAGRAM 1

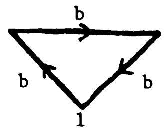

DIAGRAM 2

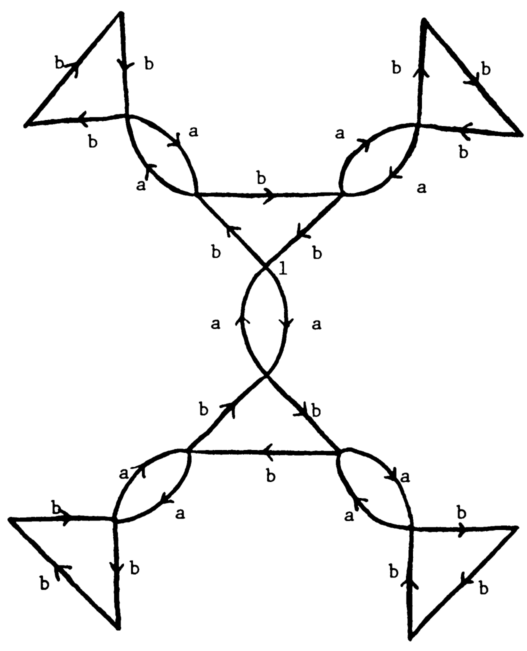

DIAGRAM 3

EXAMPLE 2. The direct product of an infinite cyclic group and a cyclic group of order two is usually presented by $\pi_2 = \langle a, b; a^2, ab = ba \rangle$. A portion of the Cayley diagram appears in Diagram 4.

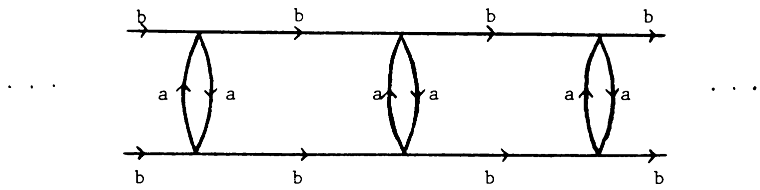

DIAGRAM 4

A *path* in a graph Δ is a sequence $p \equiv e_1 e_2 \cdots e_n$ of edges of Δ such that the initial vertex of e_{i+1} is the final vertex of e_i for $i = 1, 2, \ldots, n-1$. The *length* of the path p, written $|p|$, is n. The path is *simple* if the vertices $\alpha(e_1), \alpha(e_2), \ldots,$ and $\alpha(e_n)$ are distinct. It is *closed* if the initial vertex of e_1 and the final vertex of e_n coincide.

Just as each edge of a Cayley diagram $\Delta(\pi)$ is labelled by a generator or its inverse, each path can be labelled by a word on the generators and their inverses. Extend the labelling function μ to paths in $\Delta(\pi)$ by defining $\mu(e_1 \cdots e_n) \equiv \mu(e_1) \cdots \mu(e_n)$. Statements about words in a presentation can be reformulated as statements about paths in the corresponding Cayley diagram. The closed paths are those labelled by elements of the word problem. A closed path which returns to its initial vertex only once is labelled by an element of the reduced word problem. Words which label simple closed paths play an important role in the discussion to follow and so are given a name.

DEFINITION. Let $\pi = \langle X; R \rangle$ be a finitely generated presentation of a group G, and let $\Sigma = X \cup X^{-1}$. The set of *irreducible words* on Σ, denoted by $W(\pi)$, consists of those words equal to the identity but having no proper subword equal to the identity. □

The main result of this section establishes a less transparent correspondence between paths and words than those stated above.

THEOREM 1. *The reduced word problem of a finitely generated group presentation π is a simple language if and only if there are only a finite number of simple closed paths through each vertex of the Cayley diagram $\Delta(\pi)$.*

Theorem 1 is the geometric equivalent of Theorem 1′, which is the theorem actually proven below.

THEOREM 1′. *The reduced word problem of a finitely generated group presentation π is a simple language if and only if the set of irreducible words $W(\pi)$ is finite.*

When matching simple closed paths in the Cayley diagram with elements of $W(\pi)$, bear in mind that for each generator x, the oriented edges labelled by x and x^{-1} are distinct, although they correspond to the same geometric edge in a drawing of the Cayley diagram. Thus, if $x \neq 1$, the paths labelled by xx^{-1} and $x^{-1}x$ are simple closed paths.

The term defined next facilitates discussion of Cayley diagrams having only a finite number of simple closed paths through each vertex.

DEFINITION. Let v be a vertex of a graph Δ. The *fundamental neighborhood* of v, denoted by $\mathrm{FN}(v)$, is the subgraph of Δ consisting of all edges and vertices on some simple closed path through v. □

The Cayley diagram of a presentation is homogeneous in the sense that for any two vertices there is an automorphism of the diagram which maps one vertex to the other and maps each edge to an edge with the same label. As a result, the fundamental neighborhoods of all vertices of a Cayley diagram are isomorphic. One can reasonably speak of the fundamental neighborhood of the diagram, not simply of a particular vertex. Theorem 1 can be restated as

THEOREM 1″. *The reduced word problem* $\mathrm{WP}_0(\pi)$ *is simple if and only if the fundamental neighborhood of* $\Delta(\pi)$ *is finite.*

The Cayley diagram described in Example 1 has the finite fundamental neighborhood shown in Diagram 5. The fundamental neighborhood of the Cayley diagram described in Example 2 is the entire diagram; given any two vertices, a simple closed path passing through both vertices can be found. Hence, the presentation π_1 has a simple reduced word problem, while π_2 does not.

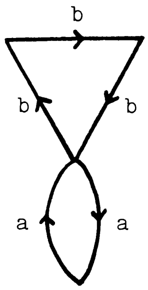

DIAGRAM 5

Now let $\pi = \langle X; R\rangle$ be a finitely generated presentation and begin the proof of Theorem 1′ by showing that if $\mathrm{WP}_0(\pi)$ is simple then $W(\pi)$ is finite. Let $\Gamma = (V, \Sigma, P, S)$ be a simple grammar generating $\mathrm{WP}_0(\pi)$. Define T to be the set of all pairs of nonempty words (u, v) on Σ such that the concatenation uv^{-1} is in $W(\pi)$. The set T is shown to be finite using the properties of simple grammars derived in §II, so $W(\pi)$ must also be finite. The next lemma lists some properties of T needed to prove it finite.

LEMMA 9. *If* $(u, v) \in T$, *then*:

(a) *neither* u *nor* v *has a nonempty subword equal to the identity*;

(b) $(v, u) \in T$;

(c) *if* $u \equiv ab$ *and* $v \equiv cd$, *then* $a = c$ *if and only if* a *and* c *are empty or* b *and* d *are empty*;

(d) *if* $u \equiv ab$ *and* a *is nonempty, then* $(a, vb^{-1}) \in T$; *and*

(e) uu^{-1} *and* uv^{-1} *are elements of* $\mathrm{WP}_0(\pi)$.

PROOF. (a) If u or v had a nonempty subword equal to the identity, then uv^{-1} would have a proper subword equal to the identity. Since $uv^{-1} \in W(\pi)$, no such subword exists.

(b) The pair (v, u) is in T if and only if vu^{-1} is in $W(\pi)$. Since vu^{-1} is the inverse of uv^{-1}, which is in $W(\pi)$, it shares with uv^{-1} the properties of being equal to the identity and having no proper subword equal to the identity. The word vu^{-1} is therefore in $W(\pi)$.

(c) If $u \equiv ab$ and $v \equiv cd$, then $uv^{-1} \equiv abd^{-1}c^{-1} \in W(\pi)$. Since $uv^{-1} = 1$, the words a and c are equal if and only if $bd^{-1} = 1$. The subword bd^{-1} of uv^{-1} is equal to the identity if and only if it is empty or it is identical to uv^{-1}. The word bd^{-1} is empty if and only if both b and d are empty, and it is identical to uv^{-1} if and only if both a and c are empty.

(d) If $u \equiv ab$, then $uv^{-1} \equiv abv^{-1} \equiv a(vb^{-1})^{-1}$. Since a and vb^{-1} are nonempty and $a(vb^{-1})^{-1}$ is in $W(\pi)$, the pair (a, vb^{-1}) is in T.

(e) The word uv^{-1} is in $W(\pi)$, a subset of $\mathrm{WP}_0(\pi)$. The word uu^{-1} is certainly equal to the identity, so it would not be in $\mathrm{WP}_0(\pi)$ only if it had a proper prefix equal to the identity. Such a prefix could have the form a or ub^{-1}, where a is a nonempty prefix of u, and b is a proper suffix of u. The former possibility can be dismissed immediately by appeal to Lemma 9(a). In the latter case, if $u \equiv cb$, then, since $u = b$, the word c is a nonempty subword of u equal to the identity. Again by Lemma 9(a), no such subword of u exists. □

LEMMA 10. *The set T is finite.*

PROOF. By Lemma 8 and the definition of T, if $(u, v) \in T$ then there is a unique leftmost derivation $S \overset{*}{\Rightarrow} u\alpha(u)$ in Γ, where $\alpha(u)$ is a nonempty string of variables. If $|\alpha(u)| \geqslant 2$ for some (u, v) in T, then $\alpha(u) \equiv A\beta$ where A is a variable and β is a nonempty string of variables. According to Lemma 9(e), both uu^{-1} and uv^{-1} are in $\mathrm{WP}_0(\pi)$. Their derivations must take the forms

$$S \overset{*}{\Rightarrow} uA\beta \overset{*}{\Rightarrow} uu_2^{-1}\beta \overset{*}{\Rightarrow} uu_2^{-1}u_1^{-1} \quad \text{and} \quad S \overset{*}{\Rightarrow} uA\beta \overset{*}{\Rightarrow} uv_2^{-1}\beta \overset{*}{\Rightarrow} uv_2^{-1}v_1^{-1}.$$

That is, the words u and v have subwords u_1 and v_1, each of which is derived from β, and u_2 and v_2, each of which is derived from A. By Lemma 3, $u_1 = v_1$ and $u_2 = v_2$. By Lemma 9(c), either u_1 and v_1 are empty or u_2 and v_2 are empty. In either case, an empty word would have been derived from a nonempty string of variables other than S. This cannot occur in a simple grammar, so $|\alpha(u)| = 1$ for all $(u, v) \in T$.

The preceding paragraph shows that $\{\alpha(u)\colon (u, v) \in T\}$ is finite by virtue of being a subset of the finite set of variables of Γ. Since $u = [\sigma(\alpha(u))]^{-1}$, the set H of elements of G equal to the first component of some element of T is also finite. By Lemma 9(d), if $(u, v) \in T$ then every nonempty prefix of u is the first component of some element of T. If the length of u, which is the number of nonempty prefixes of u, were greater than the cardinality of H then two distinct prefixes would be equal and u would have a nonempty subword equal to the identity. This cannot be, so the length of u is at most the order of H. The same bound applies to both components of each element of T since $(u, v) \in T$ if and only if $(v, u) \in T$. The set of all pairs of words on Σ satisfying this length restriction is finite, so T is finite. □

To prove the second half of Theorem 1′, that if $W(\pi)$ is finite then $WP_0(\pi)$ is simple, more must be known about the relationship between the group G presented by π and the set of subwords of $W(\pi)$. The concepts defined in the next paragraph were developed by J. R. Stallings in his study **[4]** of free products of groups. As will be seen, they are motivated by the normal form theorem for free products.

For any subset H of a group G, define a *word on* H to be a sequence $(h_1, \ldots, h_n)$ of elements of H. This word is *reduced* if no element in the sequence is the identity and no product of two consecutive elements in the sequence is itself in H. The pair (G, H) is a *unique factorization group* if the following three conditions are satisfied:

(i) The identity element is in H.

(ii) If x is in H, then so is x^{-1}.

(iii) Every element of G can be represented in exactly one way as the product of elements in a reduced word on H.

For example, let A and B be nontrivial groups, let G be their free product, and let H be the union of A and B. Certainly, H contains the identity and is closed under the taking of inverses. The normal form theorem for free products states that every nontrivial element of G has a unique expression as the product $g_1 g_2 \cdots g_n$ of elements in a normal form $(g_1, g_2, \ldots, g_n)$. Put another way, the theorem says that each nontrivial element of G has a unique representation as a product of elements in a reduced word $(g_1, \ldots, g_n)$ on H. The product of two consecutive elements in the word could not be in H without violating the uniqueness provision of the normal form theorem; both the single element $(g_i g_{i+1})$ and the pair (g_i, g_{i+1}) would be normal forms representing the same group element. The pair $(A * B, A \cup B)$ is therefore a unique factorization group. In fact, as will be shown in §IV, in any unique factorization group (G, H), the group G is a free product, though the free factors need not be so easy to identify as in $(A * B, A \cup B)$.

A key property of unique factorization groups is

LEMMA 11. *Let (G, H) be a unique factorization group. If x, y, z, xy, and yz are elements of H and y is not the identity, then xyz is in H.*

PROOF. If x, z, xy, or yz is the identity element, then xyz is surely in H. If all five elements were nontrivial and xyz were not in H, then (xy, z) and (x, yz) would be distinct reduced words for the same element of G—an impossibility. □

Henceforth, denote by H the set described below.

DEFINITION. Let π be a finitely generated presentation of a group G. Define H to be the set of all elements of G which may be represented by a subword of $W(\pi)$. □

The set H could also be described as the set of elements of G labelling vertices in the fundamental neighborhood of the identity vertex of the Cayley diagram $\Delta(\pi)$. Refer to Examples 1 and 2 above. The set H for presentation π_1 is $\{1, a, b, b^{-1}\}$; that for π_2 is the entire group.

To complete the proof of Theorem 1′, two lemmas concerning simple closed paths in a graph are proven and used to show that the property of unique factorization groups described in Lemma 11 is also held by the pair (G, H) described in the preceding definition. At that point it is shown that (G, H) is a unique factorization group itself, a fact needed in §IV. Finally, a simple grammar generating $\mathrm{WP}_0(\pi)$ is constructed from H.

It is worth pointing out that no use is made of the hypothesis that $W(\pi)$ is finite until the last step, in which the simple grammar is constructed. Thus, for all finitely generated presentations π of a group G, the pair (G, H) is a unique factorization group. This may seem a bit surprising, particularly in view of the earlier assertion that the first member of any unique factorization pair is a free product. The loophole is that H may be all of G, in which case the only nontrivial free factor of G is G itself. This is precisely the situation in the second of the recurring examples.

LEMMA 12. *Let Δ be a graph and let v_0 be a vertex of Δ. Suppose that α and β are simple closed paths through v_0 and that v and w are distinct vertices on α and β, respectively. Then there is a simple path from v to w which is composed of subpaths of α, α^{-1}, β, and/or β^{-1}, and which passes through v_0.*

PROOF. Let $\alpha \equiv ab$ and $\beta \equiv cd$, where a is a path from v_0 to v, and c is a path from v_0 to w. Both a^{-1} and b are paths from v to v_0 which intersect β (at v_0 if nowhere else). Let x and y be the first vertices of β reached by a^{-1} and b, respectively.

The vertices x and y can coincide only if both are v or both are v_0. In the former case, the vertices v and w both lie on β and so a subpath of either β or β^{-1} will satisfy the requirements of the lemma. In the latter case, the paths α and β intersect only at v_0 and there are many choices for the path sought (bc is one possibility).

If x and y are different then one of them is *not* w. Suppose, for example, that $y \neq w$ and that y lies on c. Let $c \equiv c_1c_2$, where c_1 is the subpath of c ending at y, and let $b \equiv b_1b_2$, where b_1 also ends at y. Then the path $b_1c_1^{-1}d^{-1}$ satisfies the requirements. Clearly, it goes from v to w via v_0. The paths b_1 and $c_1^{-1}d^{-1}$ are individually simple since each is properly contained in the simple circuits formed by α and β, respectively, and they intersect each other only where they should—at y—since y is the first vertex of β on b and hence the only vertex of β on b_1. Thus, $b_1c_1^{-1}d^{-1}$ is a simple path. Other cases can be handled in a similar fashion. □

LEMMA 13. *Let Δ be a graph and v_0 be a vertex of Δ. If γ is a simple path between distinct vertices in the fundamental nighborhood of v_0, then every vertex of γ is in the fundamental neighborhood of v_0.*

PROOF. Let S be the set of all paths γ such that γ is a simple path between distinct vertices in $\mathrm{FN}(v_0)$ and some vertex of γ is not in $\mathrm{FN}(v_0)$. Assume S is not empty and let γ_0 be a path in S of minimal length. Let v and w be the initial and final vertices of γ_0. No vertex of γ_0 other than v and w can be in $\mathrm{FN}(v_0)$, for if there were such a vertex u then either the subpath of γ_0 from v to u or that from u to w would be an element of S shorter than γ_0.

Since v and w are in $\mathrm{FN}(v_0)$, there are simple closed paths α and β which begin at v_0 and pass through v and w, respectively. According to Lemma 12, there is a simple path δ from v to w which passes through v_0 and is made up of subpaths of α, β, α^{-1}, and/or β^{-1}. Since all vertices of α and β are in $\mathrm{FN}(v_0)$, the same is true of δ. Thus, δ

and γ_0 intersect only at v and w, so $\gamma_0\delta^{-1}$ is a simple closed path through v. But this contradicts the fact that some vertex of γ_0 lies outside $\mathrm{FN}(v_0)$. The only conclusion possible is that S is empty. □

An analogue of Lemma 11 can now be proven.

LEMMA 14. *Let π be a finitely generated presentation of a group G and let H be the subset of G consisting of those elements equal to a subword of $W(\pi)$. If x, y, z, xy, and yz are in H and $y \neq 1$, then xyz is in H.*

PROOF. As in the proof of Lemma 11, one can assume that x, y, xy, and yz are all nontrivial. Let v_0 be the vertex of $\Delta(\pi)$ labelled by the identity. The elements x and xy label distinct vertices v and w of $\mathrm{FN}(v_0)$. Lemma 12 will be used to show that there is a simple path from v to w through the vertex labelled by xyz. By Lemma 13, each vertex on that path is in $\mathrm{FN}(v_0)$. Thus, xyz is in H.

By assumption, there are words a and b in $W(\pi)$ having proper prefixes equal to yz and z, respectively. These words correspond to simple closed paths in $\Delta(\pi)$. The path α beginning at v and labelled by a passes through the vertex v_1 labelled by xyz, as does the path β beginning at w and labelled by b. Lemma 12, applied with the vertex v_1 in the role of the lemma's v_0, implies that there is a simple path from v to w through v_1. □

Although the full strength of the next theorem is not needed until §IV, this is a good point at which to show that (G, H) is a unique factorization group.

THEOREM 3. *Let π be a finitely generated presentation of a group G and let H consist of those elements of G which are equal to a subword of $W(\pi)$. Then (G, H) is a unique factorization group.*

PROOF. The identity element of G is equal to every word in $W(\pi)$, so it is in H. The definition of $W(\pi)$ implies that $W(\pi)$ is closed under the taking of inverses and the cyclic permutation of the letters in a word. Also, if x is a generator or its inverse, then either x (if $x = 1$) or xx^{-1} (if $x \neq 1$) is an element of $W(\pi)$. The set H, therefore, is closed under the taking of inverses and contains all the generators of G. Hence, each element of G has some representation as a product of elements in a reduced word on H. Only the uniqueness of such a representation remains to be proven.

Define the length of a sequence $(w_1, \ldots, w_n)$ of subwords of $W(\pi)$ to be the sum of the lengths of the subwords. Define the length $|h|$ of a word $h = (h_1, \ldots, h_n)$ on H to be the length of the shortest corresponding sequence $(w_1, \ldots, w_n)$ of subwords of $W(\pi)$ such that $h_i = w_i$ for $i = 1, \ldots, n$.

Suppose g and h are reduced words on H representing the same element of G. Let $(g_1, \ldots, g_n)$ and $(h_1, \ldots, h_n)$ be sequences of subwords of $W(\pi)$ corresponding to g and h and having lengths $|g|$ and $|h|$. The proof shall proceed by induction on $|g| + |h|$. If the sum is zero, both words are empty and so are identical. If $|g| + |h| > 0$, then $\alpha \equiv g_1 \cdots g_n h_n^{-1} \cdots h_1^{-1}$ is a nonempty word in $\mathrm{WP}(\pi)$, which must, therefore, have a nonempty subword β in $W(\pi)$.

If β were a subword of $g_1 \cdots g_n$ alone, it could be deleted from $(g_1,\ldots,g_n)$ to obtain a shorter word representing g. Since $(g_1,\ldots,g_n)$ was chosen for its minimal length, this is not possible. For the same reason, β cannot be a subword of $h_n^{-1} \cdots h_1^{-1}$ alone. So β begins in $g_1 \cdots g_n$ and ends in $h_n^{-1} \cdots h_1^{-1}$.

Suppose β includes a nonempty portion of g_{n-1}. Let $g_{n-1} \equiv g'_{n-1}g''_{n-1}$, where g''_{n-1} is nonempty and $g''_{n-1}g_n$ is a proper subword of β. If $x = g'_{n-1}$, $y = g''_{n-1}$, and $z = g_n$, then x, y, and z satisfy the hypotheses of Lemma 14, so $xyz = g_{n-1}g_n$ is in H. But since $(g_1,\ldots,g_n)$ is reduced, the product $g_{n-1}g_n$ cannot be in H. Thus, the word β can include no portion of g_{n-1}, nor, by a similar argument, can it include any part of h_{n-1}^{-1}.

The only remaining possibility is that $\beta \equiv bd^{-1}$, where $g_n \equiv ab$, $h_n \equiv cd$, and both b and d are nonempty. There are three cases to consider:

(i) Both a and c are nonempty. If either $g_{n-1}a$ or $h_{n-1}c$ were in H, then, by an application of Lemma 14, so would $g_{n-1}g_n$ or $h_{m-1}h_m$ be. Since this would contradict the fact that g and h are reduced, the words $(g_1,\ldots,g_{n-1}, a)$ and $(h_1,\ldots,h_{m-1}, b)$ are reduced and represent the same element of G. By induction, $n = m$, $a = c$, and $g_i = h_i$ for $i = 1,\ldots,n-1$. Since $bd^{-1} = 1$, it follows that $g_n = h_n$. Thus, the words g and h are identical sequences of elements of H.

(ii) Both a and c are empty. Then $(g_1,\ldots,g_{n-1})$ and $(h_1,\ldots,h_{m-1})$ are reduced words for the same element of G, and g and h are readily deduced to be identical from the induction hypothesis.

(iii) One of a and c is empty and the other is not. If, for example, only a is empty, then $(g_1,\ldots,g_{n-1})$ and $(h_1,\ldots,h_{m-1}, c)$ are reduced words with equal products. By induction, $n-1 = m$, $g_{n-1} = c$, and $g_i = h_i$ for $i = 1,\ldots,n-2$. But then $g_{n-1}g_n = cb = cd \equiv h_m \in H$, contrary to the assumption that g is reduced. Thus, case (iii) cannot arise. □

The next lemma completes the proof of Theorem 1′.

Lemma 15. *Let $\pi = \langle X; R\rangle$ be a finitely generated presentation of a group G. If $W(\pi)$ is finite, then $\mathrm{WP}_0(\pi)$ is generated by a simple grammar.*

Proof. If $W(\pi)$ is finite, then the set H of elements of G equal to subwords of $W(\pi)$ is also finite. Let $H = \{g_0,\ldots,g_k\}$, where g_0 is the identity. Let $N = \{A_0,\ldots,A_k\}$ be a set of symbols in one-to-one correspondence with H. Let $\Sigma = X \cup X^{-1}$ and define the productions for a grammar $\Gamma = (N \cup \Sigma, \Sigma, P, A_0)$ as follows. For each $x \in \Sigma$ and $A_i \in N$,

(i) if $g_i = x$, then $A_i \to x$ is in P;

(ii) if $g_i = xg_j$ for some nonzero integer j, then $A_i \to xA_j$ is in P; and

(iii) otherwise, the production $A_i \to xA_{\tau(x)}A_i$ is in P, where $g_{\tau(x)} = x^{-1}$.

These three possibilities are both mutually exclusive and exhaustive, so for each $x \in \Sigma$ and $A_i \in N$ there is exactly one production $A_i \to x\alpha$ with $\alpha \in N^*$. In particular, the grammar Γ is simple.

It must be shown that the grammar Γ does indeed generate $\mathrm{WP}_0(\pi)$. First observe that the productions of Γ are designed to correspond to equalities in G. Since A_0 corresponds to the identity, all words derived from A_0 are equal to the identity. That is, $L(\Gamma) \subseteq \mathrm{WP}(\pi)$. Since $L(\Gamma)$ is prefix-free, to show that $\mathrm{WP}_0(\pi) = L(\Gamma)$ it suffices, by Lemma 4, to show that $\mathrm{WP}_0(\pi) \subseteq L(\Gamma)$.

The rest of the proof uses induction on the length of a word w to show that if $w \in \mathrm{WP}_0(\pi)$, then $w \in L(\Gamma)$. If $w \equiv x_1 \cdots x_r$ is in $W(\pi)$, then it can be derived from A_0 as follows: For $t = 0,\ldots,r-1$, let h_t be the element of H equal to $x_{t+1} \cdots x_r$, and let B_t be the corresponding variable. The word w has no proper subword equal to the identity, so $h_t = g_0$ if and only if $t = 0$ or $t = r$. Also, $h_t = x_{t+1}h_{t+1}$ for $t = 0,\ldots,r-1$. The productions corresponding to these equations are $B_t \to x_{t+1}B_{t+1}$, for $t = 0,\ldots,r-2$, and $B_{r-1} \to x_r$. A derivation of w can be formed by concatenating these productions:

$$B_0 \Rightarrow x_1B_1 \Rightarrow x_1x_2B_2 \Rightarrow \cdots \Rightarrow x_1 \cdots x_{r-1}B_{r-1} \Rightarrow x_1 \cdots x_r.$$

Since $h_0 = g_0$, the first variable is the start symbol and $w \in L(\Gamma)$. If w is in $\mathrm{WP}_0(\pi)$ but not in $W(\pi)$, then it has a proper subword y in $W(\pi)$. Let $w \equiv uyv$. If w has more than one subword in $W(\pi)$, assume that y is the one appearing last in w. That is, if $w = u'y'v'$ and y' is in $W(\pi)$, then u' is no longer than u. Neither u nor v can be empty, for w has no proper prefix equal to the identity. By Lemma 5, the word uv is in $\mathrm{WP}_0(\pi)$, so by induction it is in $L(\Gamma)$. Let A_i be a variable and α a variable string such that $A_0 \overset{*}{\Rightarrow} uA_i\alpha \overset{*}{\Rightarrow} uv$ represents the leftmost derivation of uv. If it can be shown that $A_i \overset{*}{\Rightarrow} yA_i$, then

$$A_0 \overset{*}{\Rightarrow} uA_i\alpha \overset{*}{\Rightarrow} uyA_i\alpha \overset{*}{\Rightarrow} uyv$$

is a derivation confirming that $w \in L(\Gamma)$.

First observe that y has no prefix which can be derived from A_i. Suppose, to the contrary, that $y \equiv y_1y_2$ and $A_i \overset{*}{\Rightarrow} y_1$. Since $A_i\alpha \overset{*}{\Rightarrow} v$, there are words v_1 and v_2 such that $v \equiv v_1v_2$ and $A_i \overset{*}{\Rightarrow} v_1$. Then $y_1 = v_1$, so $y_2v_1 = y_1^{-1}v_1 = 1$. Because $y_2v_1 \in \mathrm{WP}(\pi)$, it has a subword in $W(\pi)$. But that subword would occur later than y as a subword of w, contradicting the defining property of y. The first observation is proven.

In fact, the word y can have no prefix equal to g_i, for if an element of $W(\pi)$ has a subword e equal to g_i, then e can be derived from A_i. Since $W(\pi)$ is closed under cyclic permutation of words, the word e is a suffix of $W(\pi)$. Let $de \in W(\pi)$. Earlier in the proof of this lemma, each element of $W(\pi)$ was shown to have a derivation from A_0 in which at most one variable is present in each step. Thus, for some j, $A_0 \overset{*}{\Rightarrow} dA_j \overset{*}{\Rightarrow} de$. Since $A_j \overset{*}{\Rightarrow} e$, both g_i and g_j are equal to e. Therefore, $i = j$ and so $A_i \overset{*}{\Rightarrow} e$.

Let $y \equiv x_1 \cdots x_r$. There can be no production $A_i \to x_1$, for no prefix of y can be derived from A_i. On the other hand, the grammar Γ is constructed so that there is some production $A_i \to x_1\alpha$, $\alpha \in N^*$. Suppose $A_i \to x_1A_jA_i$ is in P, where $j = \tau(x_1)$. As shown above, there is a leftmost derivation $A_0 \overset{*}{\Rightarrow} y$, which must begin $A_0 \to x_1A_j$. Then $A_j \overset{*}{\Rightarrow} x_2 \cdots x_r$, so

$$A_i \Rightarrow x_1A_jA_i \overset{*}{\Rightarrow} x_1x_2 \cdots x_rA_i \equiv yA_i.$$

The remaining possibility is that $A_i \to x_1 B_1$ is in P for some variable B_1. Proceed by induction on j to show that there are productions $B_j \to x_{j+1} B_{j+1}$ for $j = 0, \ldots, r-1$, where $B_0 \equiv A_i$. This is true by assumption for $j = 0$. So, suppose that $0 < j < r$ and that there is a production $B_{j-1} \to x_j B_j$. Then $\sigma(B_{j-1}) = x_j \sigma(B_j)$, and $x_{j+1}^{-1}\sigma(B_j) = x_{j+1}^{-1} x_j^{-1} \sigma(B_{j-1})$. The words x_{j+1}^{-1}, x_j^{-1}, and $x_{j+1}^{-1} x_j^{-1}$ represent elements of H since each is a subword of $W(\pi)$. The elements $\sigma(B_{j-1})$ and $x_j^{-1}\sigma(B_{j-1}) = \sigma(B_j)$ are in H by construction of the grammar Γ. Lemma 14, with $x = x_{j+1}^{-1}$, $y = x_j^{-1}$, and $z = \sigma(B_{j-1})$, implies that $x_{j+1}^{-1}\sigma(B_j)$ is in H. If $x_{j+1}^{-1}\sigma(B_j) = g_k$, then $\sigma(B_j) = x_{j+1} g_k$. The element g_k cannot be the identity, for if it were, there would be a production $B_j \to x_{j+1}$ in P and thus a derivation $A_i \overset{*}{\Rightarrow} x_1 \cdots x_{j+1}$ in Γ, contrary to the fact that no prefix of y is generated by A_i. There is, therefore, a production $B_j \to x_{j+1} B_{j+1}$, in which $B_{j+1} \equiv A_k$.

The preceding paragraph shows that there is a derivation $A_i \overset{*}{\Rightarrow} y B_r$. Since productions preserve equality,

$$g_i = \sigma(A_i) = y\sigma(B_r) = 1 \cdot \sigma(B_r).$$

Hence, $B_r \equiv A_i$. □

IV. An algebraic characterization of presentations with simple reduced word problems. This section considers groups which are the free product of a finitely generated free group and finitely many finite groups. Since they are shown to be closely related to simple languages, a natural choice of name for this class of groups would be "simple groups". This term having been pre-empted by Galois, groups of the aforementioned structure will be called *plain groups*.

A convenient way to present a plain group is as follows. Let $\langle Y; \rangle$ be a free presentation of the free factor, and let $\langle Z_1; R_1 \rangle, \ldots, \langle Z_k; R_k \rangle$ be arbitrary finitely generated presentations of the finite factors. If $X = Y \cup Z_1 \cup \cdots \cup Z_k$ and $R = R_1 \cup \cdots \cup R_k$ (assuming, of course, that $Y, Z_1, \ldots, Z_k$ are pairwise disjoint), then $\pi = \langle X; R \rangle$ presents a plain group. The main point of the next theorem is that the reduced word problem of such a presentation is simple.

THEOREM 4. *Every plain group has a presentation with a simple reduced word problem, though not every presentation of a plain group need have a simple reduced word problem.*

PROOF. For the presentation π described above, the set $W(\pi)$ will be shown to be finite, from which it follows by Theorem 1′ that $\mathrm{WP}_0(\pi)$ is simple.

Let $\Sigma = X \cup X^{-1}$. Any word w on Σ can be written as a concatenation $w_1 w_2 \cdots w_n$ where each w_i is a nonempty word in the presentation of a simgle factor and consecutive words w_i, w_{i+1} come from different factors. According to the normal form theorm for free products, if w is equal to the identity, then one of the subwords, say w_i, must be equal to the identity. If w has no proper subword equal to the identity, then $w \equiv w_i \equiv w_1$.

If w is a word in the free factor, then $w \equiv xx^{-1}$ or $w \equiv x^{-1}x$ for some free generator x in Y. If w is a word in one of the finite factors, then the length of w is bounded by the order of the factor. Each factor can therefore contribute only a finite number of words to $W(\pi)$, so $W(\pi)$ is finite.

To see that some presentations of plain groups do not have a simple reduced word problem, consider a presentation having a generator x of infinite order and another nontrivial generator y which commutes with x. For each positive integer n the word $\alpha_n \equiv x^n y x^{-n} y^{-1}$ is equal to the identity. Every proper subword of α_n has one of the forms x^k, $x^k y$, $x^k y x^{-m}$, $x^k y x^{-n} y^{-1}$, y, yx^{-m}, $yx^{-n}y^{-1}$, x^{-m}, $x^{-m}y^{-1}$, or y^{-1}, where k and m are positive integers no greater than n. If one of these subwords is equal to the identity, then either a power of x is equal to the identity, y is equal to the identity, or $y = x^r$ for some nonzero integer r. Only the last of these inferences is not precluded by the hypotheses on x and y.

If $y = x^r$, then y, too, has infinite order and the preceding argument can be repeated with the roles of x and y reversed. The outcome would be that $x = y^s$ for some nonzero integer s. Then $x = y^s = x^{rs}$ and, since x has infinite order, $rs = 1$. So either $r = s = 1$ or $r = s = -1$, meaning $y = x$ or $y = x^{-1}$.

Thus, under the hypotheses on x and y, either $y = x^{\pm 1}$ or $\{\alpha_n\colon n > 0\}$ is an infinite set of words equal to the identity but having no proper subwords equal to the identity. In the latter situation, the reduced word problem of the presentation cannot be simple. So, for example, the presentation $\langle x, y;\ y = x^2 \rangle$ does not have a simple reduced word problem though it does present a plain group, namely the free group of rank one. □

The rest of this section is devoted to proving Theorem 2. Much of the development follows work by J. R. Stallings, and the reader is referred to his paper [**4**] for proofs of several lemmas.

Some new terminology will facilitate the upcoming discussion. A *partial group H* is a subset of a group G such that:

(i) the identity is in H;

(ii) if x is in H, then x^{-1} is in H; and

(iii) if x, y, z, xy, and yz are in H and y is not the identity, then xyz is in H.

Notice that any group is a partial group. A *map of partial groups* $f\colon H \to K$ is a function such that if x, y, and xy are in H, then $f(x)f(y) = f(xy)$. The *universal group* of a partial group H is a group U together with a map of partial groups $i\colon H \to U$ such that for any group A and map of partial groups $f\colon H \to A$ there is a unique homomorphism $g\colon U \to A$ such that $gi = f$. Stallings proves the following lemma.

LEMMA 16. *Every partial group has a universal group. In fact, G is the universal group of a partial group H if and only if (G, H) is a unique factorization group.*

The close relationship of partial groups and unique factorization groups is evident in their definitions. The first two conditions on H are the same in both definitions, and condition (iii) for partial groups is Lemma 11 for unique factorization groups. Furthermore, since the set H in a unique factorization pair (G, H) generates G, any map from H to a group A could be consistent with at most one homomorphism from G to A.

In determining the structure of the group G, Stallings first partitions the nontrivial elements of the partial group H into subsets, each of which, together with the identity element, forms a *component* of H. Each component is itself a partial group and contains a subgroup of G, perhaps trivial, known as the *fundamental group* for that component. From the remaining elements of the component, Stallings extracts a set (possible empty) of free generators for a free subgroup of G. His main theorem is

THEOREM 5. *If H is a partial group with a single component, then the universal group of H is the free product of the fundamental group of H and a free group freely generated by elements of H. If H is any partial group having a universal group G, then G is the free product of the universal groups of the components of H.*

Little more need be said to prove Theorem 2. If a group G has a finitely generated presentation π whose reduced word problem is simple, then a finite partial group H whose universal group is G can be constructed from the set $W(\pi)$. This is implied by Theorem 3 and Lemma 16. Since H is finite, the fundamental groups of its components are finite, the free group associated with each component has finite basis, and there are only finitely many components. Thus, by Theorem 5, the group G is the free product of a finitely generated free group and a finite number of finite groups; G is plain.

V. What next? One way to continue the work begun in this paper is to expand the classes of languages and groups considered. Simple grammars are a special type of strict deterministic grammar, which are grammars capable of generating any prefix-free deterministic context-free language. Several of the lemmas in §II can easily be generalized to deal with strict deterministic grammars, but analogues of the main theorems of §§III and IV are apparently harder to come by. A reasonable conjecture is:

> A finitely generated group G has a presentation whose reduced word problem is a strict deterministic language if and only if G is a finite extension of a plain group.

The solvability of the equivalence problem for simple grammars suggests another problem. Given a simple grammar, can one decide if it generates the reduced word problem of a group presentation, and, if so, can one determine the group involved? The answer to the second part of the question is "yes," since the grammar can be used to calculate the partial group and its components, fundamental groups, and free generating sets. The answer to the first part of the question, however, is uncertain.

REFERENCES

1. A. V. Anisimov, *Group languages*, Kibernetika **4** (1971), 18–24.
2. M. A. Harrison, *Introduction to formal language theory*, Addison-Wesley, Reading, Mass., 1978.
3. D. E. Muller and P. E. Schupp, *Groups, the theory of ends, and context-free languages*, J. Comput. System Sci. (to appear).
4. J. R. Stallings, *A remark about the description of free products of groups*, Proc. Cambridge Philos. Soc. **62** (1966), 129–134.

DEPARTMENT OF MATHEMATICS, COLLEGE OF THE HOLY CROSS, WORCESTER, MASSACHUSETTS 01610