Ergebnisse der Mathematik
und ihrer Grenzgebiete 89

*A Series of Modern Surveys in Mathematics*

Roger C. Lyndon   Paul E. Schupp

# Combinatorial Group Theory

Roger C. Lyndon Paul E. Schupp

# Combinatorial Group Theory

With 18 Figures

Springer-Verlag
Berlin Heidelberg New York 1977

Roger C. Lyndon
University of Michigan, Dept. of Mathematics,
Ann Arbor, MI 48104/U.S.A.

Paul E. Schupp
University of Illinois, Dept. of Mathematics,
Urbana, IL 61801/U.S.A.

AMS Subject Classification (1970)
Primary: 20Exx, 20E25, 20E30, 20E35, 20E40, 20Fxx, 20F05, 20F10, 20F15, 20F25, 20F55, 20H10, 20H15, 20J05
Secondary: 57A05, 55A05, 02F47, 02H15

ISBN 3-540-07642-5 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-07642-5 Springer-Verlag New York Heidelberg Berlin

Library of Congress Cataloging in Publication Data. Lyndon, Roger C. Combinatorial group theory. (Ergebnisse der Mathematik und ihrer Grenzgebiete; bd. 89). 1. Groups, Theory of. 2. Combinatorial analysis. I. Schupp, Paul E., 1937– joint author. II. Title. III. Series. QA171.L94. 512'.22. 76-12537.


Printed in Germany.

Typesetting in Japan. Printing and bookbindung: Konrad Triltsch, Würzburg.
2141/3140-543210

# Preface

The first formal development of group theory, centering around the ideas of Galois, was limited almost entirely to finite groups. The idea of an abstract infinite group is clearly embodied in the work of Cayley on the axioms for a group, but was not immediately pursued to any depth. There developed later a school of group theory, in which Schmidt was prominent, that was concerned in part with developing for infinite groups results parallel to those known for finite groups. Another strong influence on the development of group theory was the recognition, notably by Klein, of the role of groups, many of them infinite, in geometry, as well as the development of continuous groups initiated by Lie. A major stimulus to the study of infinite discontinuous groups was the development of topology: we mention particularly the work of Poincaré, Dehn, and Nielsen. This last influence is especially important in the present context since it led naturally to the study of groups presented by generators and relations.

Recent years have seen a steady increase of interest in infinite discontinuous groups, both in the systematic development of the abstract theory and in applications to other areas. The connections with topology have continued to grow. Since Novikov and Boone exhibited groups with unsolvable word problem, results in logic and decision problems have had a great influence on the subject of infinite groups, and through this connection on topology.

Important contributions to the development of the ideas initiated by Dehn were made by Magnus, who has in turn been one of the strongest influences on contemporary research. The book *Combinatorial Group Theory*, by Magnus, Karrass, and Solitar, which appeared in 1966 and immediately became the classic in its field, was dedicated to Dehn. It is our admiration for that work which has prompted us to give this book the same title. We hope that our intention has been realized of taking a further step towards a systematic and comprehensive exposition and survey of the subject.

We view the area of combinatorial group theory as adequately delineated by the book of Magnus, Karrass, and Solitar. It is not necessary for us to list here the topics we discuss, which can be seen from the table of contents. However, we would like to note that there are two broad methods running through our treatment. The first is the 'linear' cancellation method of Nielsen, which plays an important role in Chapters I and IV; this is concerned with the formal expression of an element of a group in terms of a given set of generators for the group. The second is the more geometric method, initiated by Poincaré and Dehn, which includes many of the more recent developments in 'small cancellation theory'; this method, which plays a role in Chapters II, III, and especially V, concerns the formal expression of an element of a normal subgroup N of a group G in terms of conjugates of a given set of elements whose normal closure in G is N.

We have put considerable emphasis on connections with topology, on arguments of a primitive geometric nature, and on connections with logic. In our presentation we have tried to combine a fairly self-contained exposition at a modest level with a reasonably adequate source of reference on the topics discussed. This, together with the fact that the individual chapters were written separately by the two authors, although in close collaboration, has led to considerable variation of style, which we have nonetheless sought to adapt to the subject matter.

While we do not feel it necessary to defend our inclusions, we do feel some need to justify our omissions. There are, of course, many important branches of group theory, for example, most of the theory of finite groups, that no one would claim as part of combinatorial group theory. A borderline area, with which we have made no attempt to deal here, is that of infinite groups subject to some kind of finiteness condition. Beyond these there remain a number of important topics that we believe do belong to combinatorial group theory, but which we have mentioned only briefly if at all, on the grounds that we could not hope to improve on existing excellent treatments of these topics. We list some of these topics.

1. *Commutator calculus and Lie theory.* An excellent treatment is given in Chapter 5 of the book of Magnus, Karrass, Solitar (1966). The 'Alberta notes' of Philip Hall have been republished in 1970.

2. *Varieties of groups.* The definitive work here is the book of H. Neumann (1967).

3. *Linear groups.* Treatments germaine to our topic are given by Dixon (1973) and by Wehrfritz (1973).

4. *Groups acting on trees.* This powerful method of Bass and Serre is central to our topic. An account of this theory is contained in the widely circulated notes of Serre (1968/1969), which are intended to appear in the Springer Lecture Notes series.

5. *Ends of groups.* The development of this subject by Stallings (1968, 1968, 1970, 1971) and Swan (1969) is also central to our subject; a lucid and comprehensive account, from a somewhat different point of view, is given in the book of Cohen (1972).

6. *Cohomology theory.* Of a number of excellent sources, the book of Gruenberg (1970) seems nearest to the spirit of our discussion.

We wish also to draw attention to a few other books that are especially relevant to our topic. For a history of group theory up to the early part of this century we refer to Wussing (1960). The book of Kurosh, in its various editions and translations, remains, along with the book of Magnus, Karrass, and Solitar, the classic source for information on infinite groups. The book of Coxeter and Moser (1965) contains, among other things, presentations for a large number of groups, mainly of geometric origin. We have borrowed much from the book of Zieschang, Vogt, and Coldewey (1970). On the subject of Fuchsian groups from a combinatorial point of view we recommend, in addition to the work just cited, the Dundee notes of Macbeath (1961) and the book of Magnus (1974). For an elementary exposition of the basic connections between topology and group theory we refer to Massey (1967). For a thorough discussion of decision problems in group theory we refer to Miller (1971).

# Acknowledgements

The suggestion that such a book as this be written was made in a letter from Springer Editor Peter Hilton, written from Montpellier. It is fitting that the completed manuscript should now be submitted from Montpellier.

The first author (R.C.L.) presented the first draft of some of this work in a seminar at Morehouse College, Atlanta University, in the Fall of 1969. Later versions were developed and presented in lectures at the University of Michigan and during a short visit at Queen Mary College, University of London. Much of the work was done at the Université des Sciences et Techniques du Languedoc, Montpellier, in the year 1972/73 and in the present year, and, for a shorter period, at the University of Birmingham. He is grateful for the hospitality of these universities, and also the Ruhr-Universität Bochum. He gratefully acknowledges support of the National Science Foundation (U.S.A.) and the Science Research Council (U.K.).

The second author (P.E.S.) is grateful to the University of Illinois for an appointment to the Center for Advanced Study, University of Illinois, during the academic year 1973/74. He is also grateful for the hospitality of Queen Mary College, London, and to the University of Manitoba for various periods during the preparation of this book.

We are both greatly indebted to colleagues and students, both at the universities named above and elsewhere, for discussion and criticism. For help with the manuscript we are grateful to Mme. Barrière and to Mrs. Maund. For great help and patience with editorial matters we are grateful to Dr. Alice Peters and to Roberto Minio of Springer-Verlag.

*Postscript, February 1977.* We have taken advantage of the time before going to press to bring the manuscript more up to date by adding a few new passages in the text and by substantial additions to the bibliography.

R.C.L., Montpellier 1974
P.E.S., Urbana 1974

# Notation

We have tried to use only standard notation, and list below only a few usages that might offer difficulty.

*Set theory*

$\varnothing$ is the empty set.
$X - Y$ is set difference, where $Y$ is contained in $X$.
$X + Y$ is union, where $X$ and $Y$ are disjoint.
$\{x_1, \ldots, x_n\}$ is the unordered $n$-tuple, $(x_1, \ldots, x_n)$ the ordered $n$-tuple; when there is no ambiguity we write $x_1, \ldots, x_n$ for either.
$X \subset Y$ or $X \subseteq Y$ denotes inclusion, proper or not; $X \subsetneqq Y$ denotes strict inclusion.
$|X|$ denotes the cardinal of the set $X$ (except in special contexts).

*General*

$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ denote the (non negative) natural numbers, the integers, the rationals, the reals, the complex numbers.
$\mathbb{GL}(n, K)$, $\mathbb{SL}(n, K)$, $\mathbb{PL}(n, K)$, $\mathbb{PSL}(n, K)$ denote the general, special, projective, and projective special linear group of degree $n$ over the ring $K$.

*Group theory*

1 denotes the trivial group, $\mathbb{Z}$(or $C$) the infinite cyclic group, $\mathbb{Z}_n$(or $C_n$) the cyclic group of order $n$.
$\langle U \rangle$ or $Gp(U)$ denotes the subgroup of $G$ generated by the subset $U$, and according to context, the free group with basis $U$.
$\langle X; R \rangle$, $(X; R)$, $\langle x_1, \ldots, x_n; r_1, \ldots, r_n \rangle$, as well as several other variants, denote the presentation with generaters $x \in X$ and relators $r \in R$, or the group so presented.
$H < G$ or $H \leq G$ means that $H$ is a subgroup of $G$.
$H \lhd G$ means that $H$ is a normal subgroup of $G$.
$|G|$ is the order of $G$ (finite or infinite), except in special contexts.
$|G: H|$ is the index of $H$ in $G$.

$|w|$, for $w$ an element of a free group with basis $X$, is the length of $w$ as a reduced word relative to the basis $X$.
$[h, k] = h^{-1}k^{-1}hk$ (occasionally, where indicated, $hkh^{-1}k^{-1}$).
$[H, K]$ is the subgroup generated by all $[h, k]$ for $h \in H$, $k \in K$.
$C_G(H)$, $N_G(U)$ are the centralizer and normalizer in $G$ of the subset $U$.
$G_p$ or $\mathrm{Stab}_G(p)$ is the stabilizer of $p$ under action of $G$.
Aut $G$ is the automorphism group of $G$.
$G \times H$ is the direct product.
$G * H$, $*\{G_i : i \in I\}$, or $*G_i$ denotes the free product.
$G \underset{A}{*} H$ denotes the free product of $G$ and $H$ with $A = G \cap H$ amalgamated; $\langle G, H; A = B, \phi \rangle$ denotes the free product of (disjoint) groups $G$ and $H$ with their subgroup $A$ and $B$ amalgamated according to the isomorphism $\phi: A \to B$.
$\langle G, t; t^{-1}at = \phi(a), a \in A \rangle$ denotes the indicated HNN extension of $G$.
Transformations that occur as elements of groups will ordinarily be written on the right: $x \mapsto xT$; other functions will occasionally be written on the left, e.g., $\chi(G)$ for the characteristic function of a group $G$.

## *Note on Format*

The notation (I.2.3) refers to Proposition 2.3 of Chapter I (to be found in section 2). Similarly, (I.2) refers to that section, and (I) to Chapter I.

A date accompanying a name, e.g., Smith, 1970, refers to a paper or book listed in the bibliography.

A proof begins and ends with the mark □. This mark immediately following the statement of a proposition means that no (further) proof will be given.

# Table of Contents

# Chapter I. Free Groups and Their Subgroups

## 1. Introduction

Informally, a group is free on a set of generators if no relation holds among these generators except the trivial relations that hold among any set of elements in any group. We make this precise as follows.

**Definition.** Let $X$ be a subset of a group $F$. Then $F$ is a *free group with basis* $X$ provided the following holds: if $\phi$ is any function from the set $X$ into a group $H$, then there exists a unique extension of $\phi$ to a homomorphism $\phi^*$ from $F$ into $H$.

We remark that the requirement that the extension be unique is equivalent to requiring that $X$ generate $F$.

**Proposition 1.1.** *Let $F_1$ and $F_2$ be free groups with bases $X_1$ and $X_2$. Then $F_1$ and $F_2$ are isomorphic if and only if $X_1$ and $X_2$ have the same cardinal.*

□ Suppose that $f_1$ is a one-to-one correspondence mapping $X_1$ onto $X_2$, and let $f_2 = f_1^{-1}$. The maps $f_1$ and $f_2$ determine maps $\phi_1 : X_1 \to F_2$ and $\phi_2 : X_2 \to F_1$. These have extensions to homomorphisms $\phi_1^* : F_1 \to F_2$ and $\phi_2^* : F_2 \to F_1$. Now $\phi_1^* \phi_2^* : F_1 \to F_1$ acts as the identity $f_1 f_2 = i_{X_1}$ on $X_1$, and hence is an extension of the inclusion map $X_1 \to F_1$. Since the identity $i_{F_1} : F_1 \to F_1$ also extends this inclusion map, by uniqueness we have $\phi_1^* \phi_2^* = i_{F_1}$. Similarly; $\phi_2^* \phi_1^* = i_{F_2}$. It follows that $\phi_1^*$ is an isomorphism from $F_1$ onto $F_2$.

It remains to show that $F$ determines $|X|$. The subgroup $N$ of $F$ generated by all squares of elements in $F$ is normal, and $F/N$ is an elementary abelian 2-group of rank $|X|$. (If $X$ is finite, $|F/N| = 2^{|X|}$, finite; if $|X|$ is infinite, $|F/N| = |X|$). □

**Corollary 1.2.** *All bases for a given free group $F$ have the same cardinal, the rank of $F$.* □

We remark that a free group of rank 0 is trivial.

**Proposition 1.3.** *If a group is generated by a set of $n$ of its elements ($n$ finite or infinite), then it is a quotient group of a free group of rank $n$.*

□ We assume now the existence of a free group with an arbitrary given set as basis; this will be proved below (1.7). Let $G$ be generated by the set $S \subseteq G$, $|S| = n$, let $f$ be a one-to-one correspondence from a set $X$ onto $S$, and let $F$ be free with $X$ as basis. Then $f$ determines a map $\phi: X \to G$, which extends to a homomorphism $\phi^*$: $F \to G$. Since the image $S$ of $X$ generates $G$, $\phi^*$ maps $F$ onto $G$. □

The class of free groups can be characterized without reference to bases. This results from the circumstance that, in the category of groups, projective objects are free.

**Definition.** A group $P$ is *projective* provided the following holds: if $G$ and $H$ are any groups and if $\gamma$ is a map from $G$ onto $H$ and $\pi$ a map from $P$ into $H$, then there exists a map $\phi$ from $P$ into $G$ such that $\phi\gamma = \pi$.

**Definition.** A map $\rho$ from a group $G$ onto a subgroup $S$ is a *retraction*, and $S$ is a *retract* of $G$, provided thet $\rho^2 = \rho$, or, equivalently, that the restriction of $\rho$ to $S$ is the identity on $S$.

**Proposition 1.4.** *The projective groups are precisely the retracts of free groups.*

□ Let $P$ be projective. In the definition, take $H = P$ with $\pi = i_P$, the identity on $P$, and, by (1.3), let $G$ be free with $\gamma$ from $G$ onto $P$. By the definition of projectivity, there exists $\phi: P \to G$ with $\phi\gamma = i_P$. Let $R = P\phi \leqslant G$, and let $\rho = \gamma\phi$. Then $G\rho = G\gamma\phi = P\phi = R$, and $\rho^2 = \gamma\phi\gamma\phi = \gamma i_P \phi = \gamma\phi = \rho$. Thus $\rho$ is a retraction and $R$ is a retract of $G$. Since $P\phi = R$ and $\phi\gamma = i_P$, it follows that $\phi$ is an isomorphism from $P$ onto $R$. Since $R$ is a retract of a free group, so is $P$. □

We remark that although the subgroup of a free group generated by part of a basis is obviously a retract, not every retract of a free group is of this sort; for a counterexample see Magnus, Karrass, and Solitar, p. 140.

For the following we assume, (2.11) below, that every subgroup of a free group is free.

**Corollary 1.5.** *The projective groups are precisely the free groups.* □

We turn now to the existence of free groups. This follows from principles of universal algebra (see Cohn 1965), but we prefer an explicit construction. Let a set $X$ be given; in anticipation we call the elements of $X$ *generators*. Let $Y$ be a set disjoint from $X$ with a one-to-one correspondence $\eta: X \to Y$. If $x \in X$ and $x\eta = y$ we write also $y\eta = x$ (thus $\eta$ becomes an involution on the set $X \cup Y$). We write $y = x^{-1}$ and $x = y^{-1}$, and we call $x$ and $y$ *inverse* to each other. We write $Y = X^{-1}$ and $X^{\pm 1} = X \cup X^{-1}$. The elements of $X^{\pm 1}$ are *letters*.

A *word* is a finite sequence of letters, $w = (a_1, \ldots, a_n)$, $n \geqslant 0$, all $a_i \in X^{\pm 1}$. If $n = 0$, then $w = 1$, the *empty word*. The set $W = W(X)$ of all words is a semigroup under juxtaposition (in fact, it is the free (unital) semigroup with basis $X^{\pm 1}$). With harmless ambiguity we write $a_i$ for the one-letter word $(a_i)$; this permits us to write $w = a_1 \ldots a_n$, a product of one-letter words. We extend the involution $\eta$ to $W$ by defining $w\eta = w^{-1} = a_n^{-1} \ldots a_1^{-1}$. Then $\eta$ is an involutory antiautomorphism: $(uv)^{-1} = v^{-1}u^{-1}$, $1^{-1} = 1$.

We define the *length* $|w|$ of $w = a_1 \ldots a_n$ to be $|w| = n$. Clearly $|uv| = |u| + |v|$, $|1| = 0$.

An *elementary transformation* of a word $w$ consists of inserting or deleting a part of the form $aa^{-1}$, $a \in X^{\pm 1}$. Two words $w_1$ and $w_2$ are equivalent, $w_1 \sim w_2$, if there is a chain of elementary transformations leading from $w_1$ to $w_2$. This is obviously an equivalence relation on the set $W$; moreover, it preserves the structure of $W$ as unital semigroup with involutory antiautomorphism: $u_1 \sim u_2$ and $v_1 \sim v_2$ implies that $u_1 v_1 \sim u_2 v_2$, and $u_1 \sim u_2$ implies that $u_1^{-1} \sim u_2^{-1}$. Thus we may pass to the quotient semigroup $F = W/\sim$, which is evidently a group. We shall see that it is a free group with basis the images of the $x \in X$.

A word $w$ is *reduced* if it contains no part $aa^{-1}$, $a \in X^{\pm 1}$. Let $W_0$ be the set of reduced words. We shall show that each equivalence class of words contains exactly one reduced word. It is clear that each equivalence class contains a reduced word, since successive deletion of parts $aa^{-1}$ from any word $w$ must lead to a reduced word. It will suffice then to show that distinct reduced words $u$ and $v$ are not equivalent. We suppose then that $u = w_1, w_2, \ldots, w_n = v$ is a chain from $u$ to $v$, with each $w_{i+1}$ an elementary transform of $w_i$ $(1 \leqslant i < n)$, and, indeed, with $N = \sum |w_i|$ a minimum. Since $u \neq v$ and $u$ and $v$ are reduced, we have $n > 1$, $|w_2| > |w_1|$, and $|w_{n-1}| > |w_n|$. It follows that for some $i(1 < i < n)$, $|w_i| > |w_{i-1}|, |w_{i+1}|$. Now $w_{i-1}$ is obtained from $w_i$ by deletion of a part $aa^{-1}$ and $w_{i+1}$ by deletion of a part $bb^{-1}$. If these two parts coincide, then $w_{i-1} = w_{i+1}$, contrary to the minimality of $N$. If these two parts overlap without coinciding, then $w_i$ has a part $aa^{-1}a$, and $w_{i-1}$ and $w_{i+1}$ are both obtained by replacing this part by $a$, hence again $w_{i-1} = w_{i+1}$. In the remaining case, where the two parts are disjoint, we may replace $w_i$ by the result $w'$ of deleting both parts to obtain a new chain with $N' = N - 4$, contrary to the minimality of $N$.

There is an alternative proof of the above, due to van der Waerden (1945; see also Artin 1947). For each $x \in X$ define a permutation $x\Delta$ of $W_0$ by setting $w(x\Delta) = wx$ if $wx$ is reduced and $w(x\Delta) = u$ if $w = ux^{-1}$. Let $\Pi$ be the group of permutations of $W_0$ generated by the $x\Delta$, $x \in X$. Let $\Delta^*$ be the multiplicative extension of $\Delta$ to a map $\Delta^*: W \to \Pi$. If $u_1 \sim u_2$, then $u_1\Delta^* = u_2\Delta^*$; moreover $1(u\Delta^*) = u_0$ is reduced with $u_0 \sim u$. It follows that if $u_1 \sim u_2$ with $u_1, u_2$ reduced, then $u_1 = u_2$. We note that $\Delta^*$ induces an isomorphism of $F = W/\sim$ with $\Pi$.

**Proposition 1.6.** *$F$ is a free group with basis the set $[X]$ of equivalence classes of elements from $X$, and $|[X]| = |X|$.*

□ Let $H$ be any group, and let $\phi$ map the set $[X]$ of equivalence classes $[x]$ of elements $x \in X$ into $H$. To show that $|[X]| = |X|$, we observe that if $x_1, x_2 \in X$ and $x_1 \neq x_2$, then $[x_1] \neq [x_2]$, since the two one-letter words $x_1$ and $x_2$ are reduced. Then $\phi$ determines a map $\phi_1 = X \to H$ with $[x]\phi = x\phi_1$. Define an extension $\phi_1^*$ of $\phi_1$ from $W$ into $H$ by setting $w\phi_1^* = (x_1^{e_1} \ldots x_n^{e_n})\phi_1^* = (x_1\phi_1)^{e_1} \ldots (x_n\phi_1)^{e_n}$, $x_i \in X$, $e_i = \pm 1$. If $w_1$ and $w_2$ are equivalent, then $w_1\phi_1^* = w_2\phi_1^*$, whence $\phi_1^*$ maps equivalent words onto the same element of $H$, thereby inducing a map $\phi^*: F \to H$ that is clearly a homomorphism and an extension of $\phi$. □

**Corollary 1.7.** *If $X$ is any set, there exists a free group $F$ with $X$ as basis.* □

**Proposition 1.8.** *Let $\phi$ be a homomorphism from a group $G$ onto a free group $F$ with basis $X$, and let $\phi$ map a subset $S$ of $G$ one-to-one onto $X$. Then the subgroup* Gp($S$) *of $G$ generated by $S$ is free with $S$ as basis.*

□ Let $\psi: X \to G$ be the inverse of the restriction of $\phi$ to $S$. Then $\psi$ extends to a homomorphism $\psi^*: F \to G$ with image Gp($S$). Since $\psi^*\phi$ acts as the identity on $X$, it is the identity of $F$, whence $\psi^*$ is one-to-one and so an isomorphism from $F$ onto Gp($S$) carrying $X$ onto $S$. □

**Proposition 1.9.** *Let $X$ be a subset of a group $G$ such that $X \cap X^{-1} = \emptyset$. Then $X$ is a basis for a free subgroup of $G$ if and only if no product $w = x_1 \ldots x_n$ is trivial, where $n \geqslant 1$, $x_i \in X^{\pm 1}$, and all $x_i x_{i+1} \neq 1$.*

□ Suppose first that some such $w = 1$. Let $\phi$ map $X$ injectively into a basis $Y$ for a free group $F$. Since $(x_1\phi) = \ldots (x_n\phi) \neq 1$ in $F$, $\phi$ cannot be extended to a homomorphism from Gp($X$) into $F$. It follows that $X$ is not a basis for Gp($X$).

Suppose now that no such $w = 1$. Let $F$ be a free group with a basis $Y$ in one-to-one correspondence with $X$ under $\phi: Y \to X$. Let $\phi^*$ be the unique extension of $\phi$ to a homomorphism $\phi^*: F \to G$. If $u$ is any non-trivial reduced word in $F$, then, by our hypothesis, $w = u\phi^* \neq 1$; thus $\phi^*$ is a monomorphism. Since $Y\phi^* = X$, $F\phi^* =$ Gp($X$), and $\phi^*$ is an isomorphism from $F$ onto Gp($X$) carrying $Y$ onto $X$. Since $F$ is free with basis $Y$, it follows that Gp($X$) is free with basis $X$. □

In a free group $F$ with given basis $X$, the words serve as names for the elements of $F$ in much the same way as matrices serve as names for linear transformations. Thus, if $w$ is a word, one often speaks of the group element $w$; this ambiguity must always be resolved from the context. For economy in what follows, when we speak of a free group $F$ it will be understood that $X$ is a basis for $F$, and in contrast to the usage above, the notation $|w|$ will always refer henceforth to the length of (the reduced word for) $w$ with respect to the basis $X$.

If $u$ and $v$ are elements of $F$ one has always $|uv| \leqslant |u| + |v|$; in fact, supposing $u$ and $v$ reduced one has for certain $u_1$, $v_1$, and $z$ that $u = u_1 z$, $v = z^{-1} v_1$, and $uv = u_1 v_1$ reduced. One says the parts $z$ and $z^{-1}$ have *cancelled*. It is the study of the possibilities for such cancellation in forming the product of two or more words that underlies the method of Nielsen, to which we now turn.

## 2. Nielsen's Method

The main tool in the theory of free groups, and certain related groups, is *cancellation theory*. Let $F$ be a free group with basis $X$. The words $w$ over $X$ then serve as names for the elements of $F$; in contrast to the usage above, the notation $|w|$ will denote henceforth the length of the reduced word equivalent to $w$. We say $w = u_1 \ldots u_n$ *reduced* to mean that not only does the equation $w = u_1 \ldots u_n$ hold in $F$, but also that $|w| = |u_1| + \cdots + |u_n|$; we say also that the equation holds *without cancellation* (*on the right side*). In general, given $u_1$ and $u_2$ in $F$, there exist unique $a_1$, $a_2$, and $b$ such that $u_1 = a_1 b^{-1}$, $u_2 = b a_2$, $u_1 u_2 = a_1 a_2$, all reduced, and we say that the parts $b^{-1}$ of $u_1$ and $b$ of $u_2$ have *cancelled*. Note that $|u_1 u_2| = |u_1| + |u_2| - 2|b| \leqslant |u_1| + |u_2|$. For a product of more than two elements the situation can naturally be considerably more complicated. The method of Nielsen rests upon showing that certain reasonable hypotheses limit the possibilities for such cancellation; in particular, *local* hypotheses on the amount of cancellation in a product of two or three factors lead to *global* conclusions on the amount of cancellation in a product of arbitrarily many factors.

Cancellation arguments of this sort were first applied by Nielsen to prove the Subgroup Theorem; this is very roughly related to the problem, given a subset $U$ of the free group $F$, of characterizing the elements $w$ of the subgroup $\text{Gp}(U)$ generated by $U$. An equally important problem is that of characterizing the elements of the normal closure of $U$ in $F$. Nielsen's arguments could well be called *linear* in that they deal essentially with linear arrays of symbols and transformations of them. In contrast, the second problem leads naturally to the consideration of 2-dimensional configurations, and what may be called *geometric cancellation theory*. We deal here only with the *linear* theory; the *geometric* theory will be discussed in Chapters III and V.

Nielsen first proved in 1921 by the present methods that every finitely generated subgroup of a free group is itself a free group; this is the Nielsen Subgroup Theorem. Schreier (see 3.8) using somewhat different methods proved the same conclusion without the hypothesis that the subgroup be finitely generated; this is the Nielsen-Schreier Subgroup Theorem. This more general result can also be obtained by an extension of Nielsen's method (see 2.9 below). However, the conceptually simplest proofs of these results are of a primitive topological nature (see III.3.3 below). We give a version of Nielsen's proof of his Subgroup Theorem here partly because of its elementary nature, partly because of its close analogy with a familiar argument from linear algebra, but mainly to introduce the method with a view to its many further important applications.

In considering subsets of a group $G$ it is often technically convenient to think of them as well ordered, that is to think of them as *vectors* $U = (u_1, u_2, \ldots)$, finite or infinite. However we shall not hesitate to use the same symbol $U$ for the corresponding unordered set, and indeed in many contexts we shall find it natural to deal rather with the set $U^{\pm 1}$ consisting of all $u$ and $u^{-1}$ for $u$ in $U$.

We define three types of transformation on a vector $U = (u_1, u_2, \ldots)$, as follows:

(T1) replace some $u_i$ by $u_i^{-1}$;
(T2) replace some $u_i$ by $u_i u_j$ where $j \neq i$;
(T3) delete some $u_i$ where $u_i = 1$.

In all three cases it is understood that the $u_h$ for $h \neq i$ remain unchanged. These are the *elementary Nielsen transformations*; a product of such transformations is a *Nielsen transformation*, *regular* if there is no factor of type (T3), and otherwise *singular*.

It is easy to see that each transformation of type (T1) or (T2) has an inverse which is a regular Nielsen transformation, whence the regular Nielsen transformations form a group. It is easy to see that this group contains every permutation fixing all but finitely many of the $u_i$, and also that it contains every transformation carrying $u_i$ into one of $u_i u_j$, $u_i u_j^{-1}$, $u_j u_i$, $u_j^{-1} u_i$, where $j \neq i$ (and fixing all $u_h$ for

$h \neq i$). We sometimes extend the nomenclature by counting these among the regular elementary Nielsen transformations.

**Proposition 2.1.** *If $U$ is carried into $V$ by a Nielsen transformation, then* $\mathrm{Gp}(U) = \mathrm{Gp}(V)$.

□ This is obvious for an elementary Nielsen transformation, and hence follows by induction. □

We now consider $U = (u_1, u_2, \ldots)$ where each $u_i$ is in $F$, a free group with basis $X$. As usual, $|w|$ denotes the length of the reduced word over $X$ representing $w$. We consider elements $v_1$, $v_2$, $v_3$ of the form $u_i^{\pm 1}$, and call $U$ *N-reduced* if for all such triples the following conditions hold:

(N0) $v_1 \neq 1$;
(N1) $v_1 v_2 \neq 1$ implies $|v_1 v_2| \geqslant |v_1|, |v_2|$;
(N2) $v_1 v_2 \neq 1$ and $v_2 v_3 \neq 1$ implies $|v_1 v_2 v_3| > |v_1| - |v_2| + |v_3|$.

**Proposition 2.2.** *If $U = (u_1, \ldots, u_n)$ is finite, then $U$ can be carried by a Nielsen transformation into some $V$ such that $V$ is N-reduced.*

□ Suppose first that $U$ does not satisfy (N1). Then, perhaps after a permutation of $U^{\pm 1}$, some $|u_i u_j| < |u_i|$, where $u_i u_j \neq 1$. Since it is easy to see that $|u^2| < |u|$ is impossible in a free group, we have $j \neq i$. But now a transformation (T2) replacing $u_i$ by $u_i u_j$ diminishes the sum $\sum |u_i|$. By induction we can suppose this sum reduced to its minimum, and hence that $U$ satisfies (N1). After transformations (T3) we may suppose that $U$ satisfies also (N0).

We now consider a triple $v_1 = x$, $v_2 = y$, $v_3 = z$ such that $xy \neq 1$ and $yz \neq 1$. By (N1) $|xy| \geqslant |x|$ and $|yz| \geqslant z$, whence the part of $y$ that cancels in the product $xy$ is no more than half of $y$, and likewise the part that cancels in the product $yz$. We thus have $x = ap^{-1}$, $y = pbq^{-1}$, $z = qc$, all reduced, such that $xy = abq^{-1}$ and $yz = pbc$, both reduced. If $b \neq 1$, it follows that $xyz = abc$ reduced, whence $|xyz| = |x| - |y| + |z| + |b| > |x| - |y| + |z|$, and (N2) holds for this triple. Suppose now that $b = 1$, that is, that $x = ap^{-1}$, $y = pq^{-1}$, $z = qc$, where (N2) is indeed violated. Note that we have $|p| = |q| \leqslant \frac{1}{2}|x|, \frac{1}{2}|z|$, and $p \neq q$.

In this case we have the option, by transformations of type (T2) that do not alter $\sum |u_i|$, to replace $x^{-1} = pa^{-1}$ by $(xy)^{-1} = qa^{-1}$, or to replace $z = qc$ by $yz = pc$. To avoid the situation described above we need only exercise a preference for words whose left hand halves begin with one of $p$ or $q$ over those beginning with the other. Technically, we suppose the set $X \cup X^{-1}$ of letters well-ordered. This induces a lexicographical well-ordering $u < v$ on the reduced words in $F$. We define the *left half* of a word $w$ to be the initial segment $L(w)$ of length $\left[\dfrac{|w| + 1}{2}\right]$. Finally we define a well-ordering of the pairs $\{w, w^{-1}\}$ as follows: $\{w_1, w_1^{-1}\} \prec \{w_2, w_2^{-1}\}$ if and only if either $\min\{L(w_1), L(w_1^{-1})\} < \min\{L(w_2), L(w_2^{-1})\}$ or else these two minima are equal and $\max\{L(w_1), L(w_1^{-1})\} < \max\{L(w_2), L(w_2^{-1})\}$. We shall write simply $w_1 \prec w_2$ if $\{w_1, w_1^{-1}\} \prec \{w_2, w_2^{-1}\}$. Now suppose that $x = ap^{-1}$, $y = pq^{-1}$, and $z = qc$ as above. If $p < q$ (lexicographically) then $yz = pc \prec z =$