

Ruhlen

A Guide to the World's Languages

Volume 1: Classification

A Guide to the World's Languages

Volume 1: Classification

Merritt Ruhlen



Stanford University Press 1987
Stanford, California

Stanford University Press
Stanford, California

© 1987 by the Board of Trustees of the
Leland Stanford Junior University

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Ruhlen, Merritt, 1944–
A guide to the world's languages.

Includes bibliographies and indexes.

Contents: v. 1. Classification.

1. Language and languages—Classification. I. Title.

P203.R8 1987 401'.2 86-5870

ISBN 0-8047-1250-6 (v. 1: alk. paper) Rev.

To Joseph H. Greenberg

Preface

This volume is the first in a projected series of three, which when complete will constitute a general guide to the world's languages:

Volume 1: *Classification*

Volume 2: *Language Data*

Volume 3: *Language Universals*

This first volume discusses the history and present status of the genetic classification of the world's languages and offers a complete classification. Volume 2 will provide certain kinds of linguistic and nonlinguistic information (e.g. location, number of speakers, genetic classification, consonants, vowels, pronouns, syntax) for a broad sample of the world's languages, and Volume 3 will survey the basic linguistic patterns that characterize human language around the world.

This first volume provides both a historical perspective on the development of (linguistic) genetic classification and a current assessment of the state of our knowledge. (Throughout this book I shall use the terms "genetic classification" and "linguistic classification" in place of the semantically more accurate, but stylistically awkward, "linguistic genetic classification.") Typological classification, based on structural similarities between languages, is not considered in this volume, though it will be treated in detail in Volume 3. Similarly, the historical implications of the present classification will also be explored in Volume 3.

Since human language impinges on so many related academic fields, and holds as well a general fascination for an even greater number of interested laymen, these volumes do not presume any special training on the part of the reader. In particular, this first volume does not assume any back-

ground in linguistics, genetics, or taxonomy. All that is required is a curiosity about how the world's languages are related, and about how this information has been discovered during the past two centuries. A few technical terms are introduced as needed, but the basic ideas are simple enough that the reader should encounter little difficulty.

Again with the aim of making this volume as accessible as possible to the interested layman, I have provided section-by-section topical bibliographies instead of a single comprehensive bibliography at the back of the book. By this system, the reader interested in, say, Nilo-Saharan, can easily find all the listed books and articles on that one family. Care must be taken, however, in tracing a particular citation in the text to the proper bibliography (and indeed to the proper section of that bibliography, since several bibliographies are broken into subsections).

All the linguistic maps in this volume are my own work, prepared from base maps published by the U.S. Geological Survey.

A brief sketch of the origin of this volume is perhaps in order. I first became aware of the need for a general guide to the world's languages while I was a graduate student at Stanford University from 1967 to 1973, and especially while I was working on Stanford's Language Universals Project from 1971 to 1977. During this period I compiled a *Guide to the Languages of the World* (Stanford, Calif., 1975), a self-published work containing information on roughly 700 languages. That book eventually sold some 2,500 copies, was used as a textbook at a number of universities in this country and abroad, and came to the attention of Stanford University Press, which expressed an interest in publishing a revised version.

From 1977 to 1980 I devoted my full time to collecting information on as many languages as possible, and the 7,000 data sheets I accumulated during that period contain information on about 2,000 different languages. These 2,000 languages are identified by an asterisk in the complete classification given in section 8.6, and the information itself will be found in the forthcoming Volume 2.

Since I had decided to arrange the information genetically (so that languages and groups would appear nearest the languages and groups to which they were most closely related) rather than alphabetically, as in the 1975 version (which made it easy to find individual languages, but difficult to perceive what traits characterized Romance, Germanic, Bantu, Polynesian, or any other language family), I had by late 1979 put together an overall classification of the world's languages that was based largely on C. F. Voegelin and F. M. Voegelin, *Classification and Index of the World's Languages* (New York, 1977). It was my intention to arrange the information I had collected in terms of this classification.

In November 1979 conversations with Donald Laycock and Darrell Tryon

at the Australian National University in Canberra made me aware that both the Papuan and the Austronesian classifications given in the Voegelins' book had already been largely superseded by more recent work, and they proposed that I revise my classification. They also suggested pertinent literature to consult and other scholars I might contact for advice on specific groups.

In 1980 I decided that if I was going to revise the Papuan and Austronesian sections of the classification, I should also attempt to consult experts in other language families and solicit their criticism of the classification I was then using. Correspondence with more than 100 scholars as well as a careful review of the taxonomic literature during the early 1980's led to continual refinements and revisions of the classification. Furthermore, the addition of a historical sketch and current appraisal of each family eventually metamorphosed what began as a single chapter into the present volume. The linguistic information that formed the original focus of this *Guide* will now appear in Volume 2, and Volume 3 will explore the implications of the data presented in the first two volumes.

In addition to providing a complete classification of the world's languages and a historical discussion of how, when, and by whom this knowledge was discovered, this volume has a third goal, namely, to examine the spectacular taxonomic career of the American linguist Joseph Greenberg, who has undoubtedly clarified the relationships of the world's languages more than any other scholar ever has—or ever will. Greenberg fills the same niche in linguistic classification that Linnaeus does in biological taxonomy. While other linguists were content to pursue their careers working out minute details of small, closely related language families, apparently believing that this patchwork approach would one day knit itself into an overall classification of the world's languages, Greenberg had the audacity to classify, by himself, all the languages of Africa, the languages of New Guinea, and the languages of North and South America. His rejection of established dogma in African and Amerindian classification has been met with hostility and misguided attacks by a number of traditional scholars. Nevertheless, on questions of taxonomy he has seldom been found to be wrong. It has been my rare good fortune to have known Joseph Greenberg as a teacher, colleague, and friend over a period of almost 20 years, and it is with both admiration and affection that I dedicate this book to him.

Many people have helped me in quite varied ways in writing this volume, and I wish to express my appreciation to them all. First and foremost I must thank my parents, who made everything possible, and my wife Anca, who supported my writing of this volume in ways too numerous to mention.

I must also acknowledge the extraordinary support that Stanford Univer-

sity Press has provided me during the lengthy gestation period of this volume. William Carver, in particular, has guided and encouraged my work in a multitude of ways, and his contributions have truly been substantive, organizational, and stylistic. His constant support and ability to solve any apparent problem are greatly appreciated. J. G. Bell also provided able counsel from time to time, and his advice was always taken most seriously. I would also like to thank my two fine copy editors, Kathleen Much and Barbara Mnookin, who eliminated numerous infelicities and errors from the original manuscript and rendered the final product both clearer and more readable.

A third group of people whose contributions to this volume have been enormous, and are deeply appreciated, are the more than 100 scholars (most of whom I have never met) who generously offered me advice in trying to assemble the best classification possible on the basis of current knowledge. Their suggestions have had a profound impact on the final shape of the classification, and I wish to express my sincere gratitude to them all. Their names, together with their areas of expertise, are listed on pp. xix–xx, below. I must emphasize, however, that none of these specialists necessarily agrees with the views expressed in this volume or with the final classification, for which I assume sole responsibility. Those scholars who have supplied me with their own unpublished classifications and have allowed me to use them in this book deserve special thanks: Gabriel Manessy (Gur), M. Lionel Bender (Nilo-Saharan), Robin Thelwall (Nubian and Daju), Alexander Militarev (Berber), Robert Jones (Karen), Gérard Diffloth (Austroasiatic), Lawrence Reid (Philippine languages), Paul Black (Australian), Joseph H. Greenberg (Amerind), and Eugene Loos (Panoan). I would also like to thank Sydney Lamb for reading the entire manuscript and making a number of useful suggestions, and L. L. Cavalli-Sforza, for his illuminating conversations on biological taxonomy, genetics, and related areas.

Finally I would like to thank a number of personal friends who have helped me out in various and sundry ways: Moises Moreno, Russell Ruhlén, Victor Sapojnikoff, Michael Saunders, Michael Sullivan, Andrew Wilson, and John Wilson.

M.R.

Specialists Consulted

- | | |
|---|---|
| Robert Austerlitz (Uralic, Chukchi-Kamchatkan) | James T. Collins (Maluku) |
| Yvonne Bastin (Bantu) | Bernard Comrie (Chukchi-Kamchatkan) |
| M. Lionel Bender (Nilo-Saharan, Afro-Asiatic) | *Warren Cowgill (Indo-Hittite) |
| Paul K. Benedict (Sino-Tibetan, Austic) | Irvine Davis (Ge) |
| Richard Bergman (Ubangian) | Gyula Décsy (Uralic) |
| D. N. S. Bhat (Dravidian) | Desmond Derbyshire (Carib) |
| Robert Binnick (Mongolian) | Gérard Diffloth (Austroasiatic) |
| Henrik Birnbaum (Slavic) | Søren Egerod (Sino-Tibetan) |
| Paul Black (Cushitic, Australian) | Christopher Ehret (Cushitic) |
| Robert Blust (Austronesian) | Murray Emeneau (Dravidian) |
| Nicholas Bodman (Tibeto-Burman) | Ronald E. Emmerick (Iranian) |
| Lee E. Bohnhoff (Duru) | Harold C. Fleming (Omotic) |
| Luc Bouquiaux (Adamawa, Ubangian) | Thomas Gamkrelidze (Caucasian) |
| William Bright (Hokan) | Ives Goddard (Algonquian) |
| Breandán Ó Buachalla (Celtic) | Morris Goodman (Nilo-Saharan) |
| Robbins Burling (Burmese-Lolo) | George Grace (Oceanic) |
| Wallace L. Chafe (Siouan, Caddoan, Iroquoian) | Joseph H. Greenberg (world) |
| Matthew Chen (Chinese) | Austin Hale (Sino-Tibetan) |
| Christos Clairis (Southern Andean) | Kenneth Hale (Australian, Uto-Aztecan) |
| Frans van Coetsem (Germanic) | Robert A. Hall, Jr. (Romance) |
| | Robert Harms (Uralic) |
| | André Haudricourt (New Caledonia) |

- Richard Hayward (Cushitic)
 Robert Hetzron (Afro-Asiatic)
 Carleton T. Hodge (Afro-Asiatic)
 Larry Hyman (Grassfields Bantu)
 V. V. Ivanov (Balto-Slavic)
 Philip N. Jenner (Mon-Khmer)
 Robert B. Jones (Karen)
 Terrence Kaufman (Mayan)
 Mary Key (South America)
 M. Dale Kinkade (Salish)
 Michael Krauss (Eskimo-Aleut,
 Na-Dene)
 John R. Krueger (Altaic)
 Ronald Langacker (Uto-Aztecan)
 Margaret Langdon (Hokan)
 Donald Laycock (Indo-Pacific)
 F. K. Lehman (Tibeto-Burman)
 Winfred Lehmann (Indo-European)
 Fang-kuei Li (Tai)
 W. B. Lockwood (Indo-European)
 Eugene Loos (Panoan)
 Horace G. Lunt (Slavic)
 D. N. MacKenzie (Iranian)
 Yakov Malkiel (Romance)
 Gabriel Manessy (Gur)
 Colin Masica (Indic)
 James Matisoff (Tibeto-Burman)
 David McAlpin (Dravidian)
 Curtis McFarland (Philippines)
 Alexander J. Militarev (Berber)
 Marianne Mithun (Iroquoian)
 David Moody (Borneo)
 William Moulton (Germanic)
 Paul Newman (Chadic)
 Brian Newton (Greek)
 Kemp Pallesen (Sama-Bajaw)
 F. R. Palmer (Cushitic)
 Andrew Pawley (Austronesian)
 Rebecca Posner (Romance)
 Ernst Pulgram (Italic)
 Lawrence Reid (Philippines)
 Calvin Rensch (Oto-Manguean)
 David S. Rood (Siouan)
 William J. Samarin (Adamawa,
 Ubangian)
 J. David Sapir (West Atlantic)
 Thilo C. Schadeberg (Kordofanian)
 William R. Schmalstieg (Baltic)
 Gary Simons (Santa Isabel,
 Southeast Solomons)
 William A. Smalley (Miao-Yao)
 Kenneth D. Smith (Borneo)
 David Stampe (Munda)
 Stanley Starosta (Austroasiatic)
 Susan Steele (Uto-Aztecan)
 Robin Thelwall (Nubian, Daju)
 Laurence Thompson (Salish)
 Darrell T. Tryon (Austronesian)
 Robert Underhill (Turkic)
 William S.-Y. Wang (Chinese)
 E. O. J. Westphal (Khoisan)
 Kay Williamson (Benue-Congo)
 Alfred Willms (Berber)
 Dean Worth (Chukchi-Kamchatkan)
 Norman Zide (Austroasiatic)
 David Zorc (Philippines,
 Australian)

Abbreviations

- AA *American Anthropologist*. Washington, D.C.
- AAS *Austroasiatic Studies*, 2 vols., ed. Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta. Honolulu, 1976.
- AL *Anthropological Linguistics*. Bloomington, Ind.
- AS *Afroasiatic: A Survey*, ed. Carleton T. Hodge. The Hague, 1971.
- ASEMI *Asie du Sud-Est et Monde Insulindien*. Paris.
- BAEB *Bureau of American Ethnology Bulletin*. Washington, D.C.
- CAAL *Computational Analyses of Asian and African Languages*. Tokyo.
- CSAL *Comparative Studies in Amerindian Languages*, ed. Esther Matteson. The Hague, 1972.
- CTIL *Current Trends in Linguistics*, 14 vols., ed. Thomas A. Sebeok. The Hague, 1963–76.
- EB *Encyclopaedia Britannica*, 15th ed., 1974.
- EXB *L'expansion bantoue*, 3 vols., ed. Larry M. Hyman, Jan Voorhoeve, and Luc Bouquiaux. Paris, 1980.
- GR *Genetic Relationship, Diffusion and Typological Similarities of East & Southeast Asian Languages*, ed. Mantaro J. Hashimoto. Tokyo, 1976.
- HNAI *Handbook of North American Indians*, 20 vols., ed. William C. Sturtevant. Washington, D.C., 1978– .
- IEL *Inventaire des études linguistiques sur les pays d'Afrique noire d'expression française et sur Madagascar*, ed. Daniel Barreteau. Paris, 1978.
- IJAL *International Journal of American Linguistics*. Chicago.
- LNA *The Languages of Native America*, ed. Lyle Campbell and Marianne Mithun. Austin, Tex., 1979.

- LSNA *Linguistic Structures of Native America*, ed. C. Osgood. New York, 1946.
- MKS *Mon-Khmer Studies*. Honolulu.
- NLA *Native Languages of the Americas*, 2 vols., ed. Thomas A. Sebeok. New York, 1976–77.
- NS *Nilo-Saharan*, ed. Thilo C. Schadeberg and M. Lionel Bender. Dordrecht, 1981.
- NSLE *The Non-Semitic Languages of Ethiopia*, ed. M. Lionel Bender. East Lansing, Mich., 1976.
- OL *Oceanic Linguistics*. Honolulu.
- PL *Pacific Linguistics*. Canberra.
- PLNGLS *Papuan Languages and the New Guinea Linguistic Scene*, ed. S. A. Wurm. 1975. PL C38.
- UCPAAE *University of California Publications in American Archaeology and Ethnology*. Berkeley.
- UCPL *University of California Publications in Linguistics*. Berkeley.

A Note on Notation

Although this volume does not employ a great deal of esoteric notation, some of its conventions and symbols may be unfamiliar to the general reader and so require a few words of explanation.

1. A dagger (†) identifies those languages, or language families, that are now extinct.

2. An asterisk (*) is used in two ways. Usually it identifies reconstructed (but historically unattested) words, e.g. Indo-European **akva-s* 'horse.' The second use of the asterisk is restricted to Chapter 8, where it identifies those languages whose sound system will in due course be found in Volume 2, e.g. **latmul*.

3. A right arrow (→) is used to indicate that one language (or family) name has been replaced by another. For example, Barea → Nera means that the language frequently referred to as Barea is called Nera in this volume.

4. An equals sign (=) is used to point out alternate language (or family) names that for reasons discussed in section 8.1 are sometimes better known than the name adopted here, e.g. Nama (= Hottentot).

5. Brackets ([. . .]) enclose phonetic transcriptions, e.g. [dori] 'to wish for, desire.' In Chapter 8 the number of languages in any group is indicated in brackets, following the name of the group; e.g. CELTIC [4] means there are four extant Celtic languages.

6. Single quotes ('. . .') enclose English glosses of foreign words, e.g. French *chien* 'dog.'

7. Double quotes (" . . .") are used to indicate direct quotation from a source or to call attention to a word or phrase, e.g. there are no "primitive" languages.

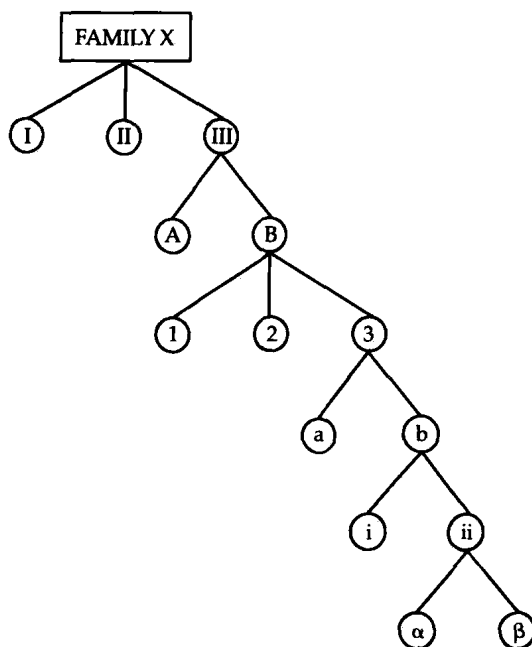
8. Italics are used for non-English words (e.g. French *chien* 'dog'), for individual letters in running text (treated as phonetic elements, whether in English or not), and for simple emphasis (e.g. *typological* traits like gender).

9. Names of language groups are given in capital letters (e.g. SLAVIC, ROMANCE) in the lists and in the classification itself; names of single languages are in capital-and-lower-case letters (e.g. Russian, English, Ohlone). In the case of single languages that are themselves independent branches of a family, either notation may be used, e.g. ALBANIAN or Albanian.

10. The less-than sign (<) indicates that a word, sound, or meaning derives historically from another, e.g. French *chien* < Latin *cane(m)*; French *š* < Latin *k*.

11. The greater-than sign (>) indicates that a word, sound, or meaning has changed into another over the course of time, e.g. Latin *cane(m)* > French *chien*; Latin *k* > French *š*.

12. The similar-to sign (~) indicates either (a) that certain forms alternate morphologically in a language, e.g. English *sing* ~ *sang* ~ *sung*, Indo-European **ped-* ~ **pod-* 'foot,' or (b) that variant forms of a root are found in different languages of a language family, e.g. Nilo-Saharan *tok* ~ *tek* ~ *dik* 'one.'



FAMILY X:
 I PRIMARY BRANCH
 II PRIMARY BRANCH
 III PRIMARY BRANCH:
 A GROUP
 B GROUP:
 1 SUBGROUP
 2 SUBGROUP
 3 SUBGROUP:
 a BRANCH
 b BRANCH:
 i SUBBRANCH
 ii SUBBRANCH:
 α GROUP
 β GROUP

13. A hyphen (-) is sometimes used to separate a root from its affixes, either prefixes (e.g. English *pre-fix*) or suffixes (e.g. Greek *dídō-mi* 'I give'). It is also used with prefixes and suffixes by themselves (e.g. the English prefix *pre-*, the Greek suffix *-mi* 'I') or following a root (e.g. Indo-European **ped-* ~ **pod-* 'foot').

14. The system of representing genetic tree structure is explained in section 8.2, but because such trees occur in lists throughout the book, a brief outline of the notation is given here. Genetic classifications are often represented in terms of branching diagrams, as in the figure on the facing page.

In this book such structures are represented by a system of indentation, where each successive level of branching is indicated by indentation to the right and each node is preceded by a letter or number to identify the taxonomic level of the group. Accordingly, the branching-tree diagram in the figure will have the representation shown in the list above.

A Guide to the World's Languages

VOLUME 1: CLASSIFICATION

Contents

Specialists Consulted xix

Abbreviations xxi

A Note on Notation xxiii

Introduction 1

1 Genetic Classification: Principles and Methods 4

1.1 Definitions and Methods 4

1.2 Continuous and Discrete Variation 6

1.3 Genetic Affinity 9

1.4 Subgrouping 14

1.5 Taxon Names 20

2 Europe 24

2.1 The Historical Context 25

2.2 The Linguistic Background 28

2.3 Indo-Hittite 35

2.4 Uralic-Yukaghir 64

2.5 Caucasian 71

3 Africa 76

3.1 The History of African Classification 76

3.2 Afro-Asiatic 85

3.3 Niger-Kordofanian 95

3.4 Nilo-Saharan 107

3.5 Khoisan 114

3.6 Criticism of Greenberg's Methodology 120