Acknowledgments

The production of *The Unicode Standard, Version 3.0*, is due to the dedication of many individuals for over a decade. We would like to acknowledge those who were involved in Version 3.0, and particularly the following individuals, whose major contributions were central to the design, authorship, and review of this book.

Ken Whistler has been the driving force behind Version 3.0. As managing editor, he had responsibility for all aspects of production; in particular, he designed the new chapter structure. Ken meticulously updated the Unicode Character Database, including adding all the new characters and their properties. He also maintained the Character Names List and supplied many of the annotations. Ken co-authored three Unicode Technical Reports, and contributed to the descriptions for many of the new scripts, such as Myanmar and Khmer. He was responsible for the vast majority of the original character block descriptions and made major contributions to the encoding of Indic scripts, IPA, and symbols.

Julie D. Allen coordinated the editorial review meetings. As editor, she kept track of edits and text from multiple sources and in various formats, and formatted the text in FrameMaker. She worked with the authors to improve the wording and clarity of the text, and also helped resolve font problems resulting from platform conversion.

Joan Aliprand chaired the Unicode Technical Committee while Version 3.0 was under development. She is responsible for the front matter, *Glossary*, *References*, and *General Index*. She also contributed to *Middle Eastern Scripts.*

Joe Becker is acknowledged for his initial work and promotion of a 16-bit full encoding and his overall contributions to the creation of the Unicode Standard and editing of this book.

Mark Davis has been essential to the development of Version 3.0. Mark has led the way in many aspects of overall design and implementation of the Unicode Standard. He has been the chief architect for the definition of conformance and the bidirectional algorithm. He contributed significant revisions and enhancements to *Implementation Guidelines*, *Bidirectional Behavior, Syriac*, and the Unicode Character Database. Five of the Unicode Technical Reports are Mark's work, and he co-wrote two others. His previous contributions include algorithms for canonical equivalence, conjoining Hangul, shaping for specific scripts, and the development of the Unicode Character Database.

Michael Everson collected or created many of the fonts used in this standard, and provided typographical review and assistance. He also reviewed the code charts and Character Names List, and provided many annotations. Block descriptions for several of the new scripts are based on drafts prepared by Michael.

Asmus Freytag was responsible for the chapters *Punctuation*, *Symbols*, and *Special Areas and Format Characters.* He made significant contributions to *Syriac*, *Bidirectional Behavior*, and to the chapters *General Structure*, *European Alphabetic Scripts*, and *Implementation Guidelines.* Asmus created custom formatting software, negotiated with the font houses for donated fonts, and oversaw production of the code charts. He also created the Shift-JIS Index, and he wrote two Unicode Technical Reports and co-authored another. Previously, Asmus contributed substantial text on implementation aspects of the Unicode Standard.

John H. Jenkins was responsible for *East Asian Scripts* and contributed substantial new material for that chapter. He maintained and extended the Unihan database, and created the Han Radical-Stroke Index to the ideographic content of the standard. John converted all of the Version 2.0 illustrations to Adobe Illustrator format. He created the fonts for the KangXi and CJK radicals. In addition, he collected fonts from many sources and fine-tuned them for use. John was responsible for the original Han cross-reference tables.

Mike Ksar led the effort to synchronize Version 3.0 and the second edition of ISO/IEC 10646-1. He contributed to *Middle Eastern Scripts*, as well as to *Appendix C* and *Appendix D*. Mike also thoroughly reviewed the code charts and the text.

Rick McGowan coordinated work on new scripts, and was involved with many of the substantial additions published in Version 3.0. He organized a group of Tibetan experts who developed the extensions for Tibetan script. Rick contributed to the sections on *Ogham*, *Thaana*, *Tibetan*, *and Canadian Aboriginal Syllabics*. He was responsible for mastering and producing the CD-ROM. Rick contributed to development of the original algorithm for canonical equivalence.

Lisa Moore rewrote many parts of the text, including the *Preface* and *Introduction*, the chapter *General Structure*, and *Appendix D*. She wrote the section on *Cherokee*, and contributed to the explanations on encoding forms and transformation formats. As current Chair of the Unicode Technical Committee, Lisa oversaw the finalization of the content of Version 3.0. She previously contributed substantially to character properties.

Michel Suignard painstakingly recoded and converted fonts for use in this book. He also contributed to several Unicode Technical Reports. Michel previously contributed to the development of the Unicode Character Database and mapping tables.

Fonts were essential for the production of this book. In addition to the individuals mentioned above and the companies and organizations named in the colophon, fonts were contributed by Michael Stone (Armenian), Al Webster (Cherokee), Oliver Corff (Yi), Yannis Haralambous (Syriac), and George Kiraz and Paul Nelson (Syriac). John M. Fiscella designed the fonts for symbols and many of the alphabetic scripts. Thomas Milo designed the Arabic font. The font for CJK Extension A was designed by Technical Supervisor Zheng Long and Hua Weicang. Cora Chang created the font for Braille patterns. Critical comments and work on fonts are due to Heidi Jenkins, and Dr. Virach Sornlertlamvanich (for Thai).

Ralph Buchalter transferred the Version 2.0 master files to the Windows environment, and implemented the new chapter structure. Phil Huston provided FrameMaker expertise. Steve Mehallo designed the interchapter separators, incorporating examples of writing from numerous sources. Kamal Mansour (Monotype) initiated and coordinated this graphic design process.

Both the text and the code charts were reviewed critically by experts. Edwin Hart and Bruce Paterson were principal reviewers of this book. The Editorial Committee appreciates the expert feedback on the code charts provided by Keiko Nagano and Tetsuji Orita of IBM Japan, Hugh McGregor Ross, and, for specific scripts, Yannis Haralambous (Greek and Thai), Constantine Stathopoulos (Greek), and Dr. Virach Sornlertlamvanich (Thai). Barbara Beeton of the American Mathematical Society reviewed the math symbols, and Donald E. Knuth gave additional input. Cora Chang thoroughly checked the data for CJK ideographs. Mark Davis and Laura Werner did extensive work to verify the consistency of character properties in the Unicode Character Database. Shen Fucheng of RLG helped with the bibliography.

The development of this book would not have been possible without the support of the office staff of Unicode, Inc., and the hard work of Mike Kernaghan, as operational manager of the office. We thank Magda Danish, Dee Domingo, Steve Greenfield, Cheryl Morton,

and Julia Oesterle, all of whom assisted with Version 3.0 in numerous ways. Sarasvati minded the mailing lists.

The technical content of the Unicode Standard is determined by the Unicode Technical Committee. Contributors to the work of the UTC include representatives of Full and Associate Members, Specialist and Individual Members, and Unicode Officers, as well as invited experts and liaisons. Version 3.0 would not have been possible without their creative work and critical thinking over the past three years, augmented by the support of all member representatives: Dean Abramson (Gamma), Glenn Adams (Spyglass), Joan Aliprand (RLG), Joshua Arai (Justsystem), Joe Becker (Xerox), Barbara Beeton (American Mathematical Society), Chris Boyle (Novell), Stefan Buchta (Oracle), Richard C. Buhler (Novell), Don Carroll (Hewlett-Packard), Steve Cohen (Basis Technology), Lee Collins (Apple), Ken Cortese (Booz, Allen), David O. Craig (NCR), Frank da Cruz (Columbia University), Mark Davis (Unicode President), Martin Dürst (W3C), Peter Edberg (Apple), John M. Fiscella (Production First), Asmus Freytag (Unicode Vice-President), Deborah Goldsmith (Apple), Edwin F. Hart (SHARE), Hideki Hiura (Sun), Paul Hoffman (Internet Mail Consortium), Lloyd Honomichl (Novell), John Jenkins (Apple), Nailong Jiang (Quark), Robert Keeble (Quark), Susan Kline (Hewlett-Packard), Tatsuo Kobayashi (Justsystem), Mike Ksar (Hewlett-Packard), Michael Kung (Oracle, then Microsoft), Wai Man Long (Compaq), Kamal Mansour (Monotype), Benson Margulies (Basis Technology), Gianni Mariani (SGI), Rick McGowan (NeXT, then Apple), Mike McKenna (Sybase), Thomas Milo (DecoType), Matthias Mittelstein (SAP), Lisa Moore (IBM), Nobuyoshi Mori (SAP), Rudolf Munz (SAP), Brendan Murray (Lotus), Nelson Ng (Oracle), Thomas Nielsen (MGI), Cleo O'Brien (SGI), Sandra Martin O'Donnell (Compaq), Stephen P. Oksala (Unisys), Shripad Patki (Sun), Nathan Potechin (MGI), Mirko Raner (MATHEMA), Wendy Rannenberg (Compaq), Gary Roberts (NCR), Murray Sargent (Microsoft), Bernhard Schilling (SAP), Karen Smith-Yoshimura (RLG), Neil Soiffer (Wolfram Research), Mark Son-Bell (Sun), Michel Suignard (Microsoft), Tex Texin (Progress Software), V. S. Umamaheswaran (IBM), Marc von Holzen (SGI), Ken Whistler (Sybase), Michael Wiedeking (MATHEMA), Richard A. Willis (Reuters), Arnold F. Winkler (Unisys), Bill Wolf (Booz, Allen), Misha Wolf (Reuters), Eugene Wong (Hewlett-Packard), Jianping Yang (Oracle), Michael Yau (Oracle), Douglas-Val Ziegler (Xerox).

A significant part of the UTC's work was devoted to refining the Unicode Bidirectional Algorithm. The task was carried out by the Bidirectional Subcommittee, consisting of Asmus Freytag (ASMUS, Inc.), Mike Ksar (Hewlett-Packard), Mati Allouche, Mark Davis, Doug Felt, Israel Gidali, Lisa Moore (all from IBM), F. Avery Bishop, David Brown, John McConnell, Paul Nelson (all from Microsoft), and Jony Rosenne (QSM Programming). The reference implementations on the CD-ROM were developed by Doug Felt and Asmus Freytag.

The Arabic Ad Hoc group reviewed the existing Arabic script repertoire and decompositions. Members of the Ad Hoc group were as follows: Joe Becker (Xerox), Asmus Freytag (ASMUS, Inc.), Mike Ksar (Hewlett-Packard), Kamal Mansour (Monotype), Thomas Milo (DecoType), Khaled Sherif (IBM), Ken Whistler (Sybase).

The Unicode Technical Committee has worked closely with Technical Committee L2 of the National Committee for Information Technology Standards (NCITS). We appreciate the cooperation of Chair Arnold Winkler and Vice-Chairs Joan Aliprand and Lisa Moore. Arnold Winkler established and efficiently maintained the Web-based archive of L2 documents that was crucial to the development of Version 3.0.

The growth of the synchronized character repertoires of the Unicode Standard and International Standard ISO/IEC 10646 reflects a worldwide effort conducted over many years. With Version 3.0, the Unicode Standard encodes all of the major modern scripts of the

world. We express deep appreciation to the following experts who shared their specialized knowledge to bring about this achievement:

- For Runic: the ISORUNES project of the Swedish Central Board of National Antiquities (Helmer Gustavson, leader, Olle Järnefors, James Knirk, Svante Lagman, Ray Page, Marie Stoklund, and Klaus Düwel).
- For Ogham: Michael Everson and Marion Gunn.
- For Syriac: George Kiraz and Paul Nelson, who provided the draft block introduction for Syriac with the assistance of their colleague Sargon Hasso. Also those who advised them: Dr. Sebastian Brock (Oxford University), Dr. J. Coakley (Harvard University), Dr. Jan Wilson (Brigham Young University), and Dr. Edward Odisho (North Eastern Illinois University).
- For Thaana: the Maldivian Students' Association and Husine Zahid.
- For Sinhala: Lloyd Anderson, Lee Collins, Camillus Jayewardena (University of Ruhuna, Matara, Sri Lanka), Vasantha Karunaratne, Hugh McGregor Ross, and Mettavihari (IBRIC—International Buddhist Research & Information Centre, Sri Lanka).
- For Tibetan extensions: Lloyd Anderson, Robert Chilton, Michael Everson, Christopher Fynn, Peter Lofting, Sam Sirlin, Valeriy E. Ushakov, and especially Tony Duff, who drafted the block introduction for Tibetan.
- For Myanmar: the Myanmar Information Technology Standardization Committee (represented by Dr. Aung Maw, Khin Maung Lwin, Dr. Kyaw Thien, Thaung Tin, and Thien Htut), and also F. Avery Bishop, Lee Collins, Andy Daniels, Zaw Htut, and Hugh McGregor Ross.
- For Khmer: Maurice Bauhahn, F. Avery Bishop, Andy Daniels, Anshuman Pandey, and Cathy Wissink.
- For Ethiopic: the Committee for the Standardization of Ethiopian Script (Abass Alamnehe, Fesseha Atlaw, Yitna Firdyiwek, Yonas Fisseha, Samuel Kinde, Terrefe Ras-Work, and especially Daniel Yacob).
- For Cherokee: Charles Gourd of the Cherokee Nation of Oklahoma.
- For Canadian Aboriginal Syllabics: the Canadian Aboriginal Syllabic Encoding Committee, as well as Alain LaBonté, Henry Leperlier, and especially Dirk Vermeulen.
- For Mongolian: Chen Zhuang, Professor Choijingzhab, Oliver Corff, Myatav Erdenchimeg, Ji Mu Yan, Yumbayariin Namsarai, and, in particular, Richard Moore, who drafted the block introduction for Mongolian.
- For the ideographs of Vertical Extension A, the Kangxi Radicals, and the CJK Radicals Supplement: all of the members of the IRG. Major contributors included Zhang Zhoucai (rapporteur), Wang Xiaoming, Chen Zhuang, Fu Yonghe, Bao Huiyun, Zheng Long, Liu Jiayu; Takayuki Sato, Satoshi Yamamoto, Tateo Koike, Eiji Matsuoka, Tatsuo Kobayashi; Lee Joonsuk, Lee Jaehoon, Suh Kyung Ho; Ngo Trung Viet, Hong Nguyen Quang; Lu Qin, Lee Kin Hong; C. C. Hsu, Emily Yu-Chi Hsu, T. C. Kao; Michael Kung, Hideki Hiura, and John Jenkins.
- For advice on Hebrew: Dr. Bella Hass Weinberg.
- For advice on Indic scripts: James Agenbroad, Jeroen Hellingman, and Antoine Lecault.

- For assistance with character properties of Thai: Ranat Thopunya and Suwit Srivilairith (IBM Thailand).
- For contributions to the encoding of North American scripts: Eric Brunner.

The Unicode Standard, Version 3.0, would not have been possible without those who made important contributions to earlier versions: Glenn Adams for his work on the character/glyph model, Devanagari shaping behavior, and the definition of conformance, and for the initial glossary; Lee Collins, for his painstaking work in the integration and extension of mappings from many different sources to create the original Unicode Han unification; Edwin Hart for his work on the character/glyph model; Isai Scheinberg, for his pivotal work in guiding the Unicode Standard through the intricacies of the ISO standards development process and his contributions to the bidirectional algorithm; and Layne Cannon and Stuart Smith for the Byte Order Mark. Past contributors to the development of the character databases and mapping tables include K. D. Chang, Tim Greenwood, Lori Hoerth, and Liao Huan-Mei. Other significant contributors to earlier versions were John Bennett, Joe Bosurgi, Andy Daniels, Burwell Davis, Bill English, Masami Hasegawa, Greg Hullander, Mike Kernaghan, Eric Mader, Dave Opstad, Bill Tuthill, J. G. Van Stee, and Glenn Wright.

The Unicode Consortium continues to maintain mutually beneficial relationships with international standards organizations. We appreciate the efforts and support of the members of ISO/IEC JTC1/SC2/WG2 and the members of the Ideographic Rapporteur Group toward the common goal of keeping both standards synchronized. We would particularly like to thank the Convener of WG2, Mike Ksar, the Rapporteur of the IRG, Zhang Zhoucai, and WG2's Editors and Contributing Editors, especially Bruce Paterson, Michel Suignard, and Michael Everson. We also thank Asmus Freytag for his effective representation of the Unicode Consortium at WG2 meetings. We would like to thank the members of ISO/IEC JTC1/SC22/WG20, especially Alain LaBonté and the Convener, Arnold F. Winkler, for their work with the Consortium on a common collation algorithm and properties related to identifiers.

The support of member companies has been crucial to *The Unicode Standard, Version 3.0*. Particular thanks for facilities, equipment, and resources are owed to Adobe Systems, Inc., Apple Computer, Inc., Hewlett-Packard Company, IBM Corporation, Microsoft Corporation, and Monotype Typography, Inc.

While we gratefully acknowledge the contributions of all persons named in this section, any errors or omissions in this work are the responsibility of the Unicode Consortium.

Unicode Consortium Members and Directors

Full Members

While Version 3.0 of the Unicode Standard was under development, the following companies were Full Members of the Unicode Consortium:

Apple Computer, Inc.
Basis Technology Corporation
Booz, Allen & Hamilton, Inc.
Compaq Computer Corporation
Gamma Productions, Inc.
Hewlett-Packard Company
IBM Corporation
Justsystem Corporation
MATHEMA Software GmbH
MGI Software Corporation
Microsoft Corporation
NCR Corporation
NeXT Software, Inc.
Novell, Inc.
Oracle Corporation
The Research Libraries Group, Inc.
Reuters, Ltd.
SAP AG
Silicon Graphics, Inc.
Spyglass, Inc.
Sun Microsystems, Inc.
Sybase, Inc.
Unisys Corporation
Xerox Corporation

Current Associate Members

Adobe Systems, Inc.
Arphic Technology Co. Ltd.
ASMUS, Inc.
Automated Wagering International, Inc.
Beijing Zhong Yi Electronics Co.
Bitstream, Inc.
BMC Software, Inc.
The Church of Jesus Christ of Latter-day Saints
Columbia University
Data Research Associates
DecoType, Inc.
Endeavor Information Systems, Inc.
Ericsson Mobile Communications
eTranslate, Inc.
GE Information Services, Inc.
The Getty Information Institute
Global System for Sustainable Development
Government of Tamil Nadu, India
Hong Kong Telecom CSL
Innovative Interfaces, Inc.
Internet Mail Consortium
Language Analysis Systems, Inc.
Libronix Corporation
Logos Research Systems, Inc.
Lycos, Inc.
Monotype Typography, Inc.
Netscape Communications Corporation
OCLC, Inc.
OneRealm, Inc.
Production First Software
Progress Software Corporation
QUALCOMM Incorporated
Quark, Inc.
Rogue Wave Software
The Royal Library, Sweden
RWS Group, LLC
SHARE, Inc.
Siebel Systems, Inc.
Software AG
Software Engineering Australia, Ltd.
StarTV—Satellite Television Asia Region Ltd.
Symbian, Ltd.
Uniscape, Inc.
VTLS, Inc.

Current Liaison Members

Association for Font Information Interchange (AFII)
China, Center of Computer and Information Development (CCID), Beijing
The Internet Engineering Task Force (IETF)
Korea, Kongju National Library, Chung-nam
Viet Nam, Technical Committee on Information Technology (TCVN/TC1), Hanoi
World Wide Web Consortium (W3C) I18N Working Group

Current Specialist Members

James Agenbroad, Rolf Bourges, G. David Butler II, James Caldwell, Peter Constable, Tony Graham, Stephen Holmes, Kent Karlsson

Current Individual Members

Lloyd Anderson, Stephanos Antoniades, Dr. Yves Arrouye, Stella Ayzinger, Hasan Bateson, Charles Bishop, Philip Blair, Brad Bodkin, François-Paul Briand, Brian M. Carr, Shang-Che Cheng, Marco Cimarosti, John Dlugosz, George Foot, Richard A. Gard, Dr. Harry Gaylord, Dan Glisson, Jennifer Goodman, Tim Greenwood, Phillip H. Griffin, Bob Hallisy, James Harlow, Nyr Indictor, Jens K. Jensen, Bohdan Kantor, James Kass, Tim Kimball, W. J. Kurmey, Rob Latousek, Kevin Leake, Jonathan Lewis, Giovanni Lussu, Pankaj Madia, Dr. Karnati Mallesham, Greg Melahn, Dr. Tag Young Moon, Robert J. Morris, Victor Nguyen, Paul Pavlik, Phan Phan, Gabriel Plumlee, Christopher Pullen, Patrick P. Reilly, Daniel Rybowski, Tetsuo Seto, CathyAnn Swindlehurst, Lang-Fen Hu Tang, Mituhiko Toho, Tom Twiddy, L. Segond von Banchet, Joan Winters

Current Members of the Board of Directors

Wayne R. Boyle (NCR Corporation)
Brian E. Carpenter (IBM Corporation & Internet Architecture Board)
Kevin Cavanaugh (Lotus Development Corporation)
Tatsuo L. Kobayashi (Scholex Company, Ltd.)
Mike Ksar (Hewlett-Packard Company)
Elizabeth G. Nichols (IBM Corporation)
Wendy Rannenberg (Compaq Computer Corporation)
Franz G. Rau (Microsoft Corporation)
David R. Richards (The Research Libraries Group, Inc.)

Former Members of the Board of Directors

Jerry Barber (Aldus Corporation)
Janet Buschert (Hewlett-Packard Company)
Robert M. Carr (Go Corporation)
John Gage (Sun Microsystems, Inc.)
Paul Hegarty (NeXT Software, Inc.)
Gary C. Hendrix (Symantec Corporation)
Richard J. Holleman (IBM Corporation)
Charles Irby (Metaphor, Inc.)
Jay E. Israel (Novell, Inc.)
Ilene H. Lang (Digital Equipment Corporation)
Paul Maritz (Microsoft Corporation)
Stephen P. Oksala (Unisys Corporation)
Mike Potel (Taligent, Inc.)
Rick Spitz (Apple Computer, Inc.)
Lawrence Tesler (Apple Computer, Inc.)
Guy “Bud” Tribble (NeXT Software, Inc.)
Kazuya Watanabe (Novell KK)
Gayn B. Winters (Digital Equipment Corporation)

This PDF file is an excerpt from *The Unicode Standard, Version 3.0*, issued by the Unicode Consortium and published by Addison-Wesley. The material has been modified slightly for this online edition, however the PDF files have not been modified to reflect the corrections found on the Updates and Errata page (see http://www.unicode.org/unicode/uni2errata/UnicodeErrata.html). More recent versions of the Unicode standard exist (see http://www.unicode.org/unicode/standard/versions/).

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and Addison-Wesley was aware of a trademark claim, the designations have been printed in initial capital letters. However, not all words in initial capital letters are trademark designations.

The authors and publisher have taken care in preparation of this book, but make no expressed or implied warranty of any kind and assume no responsibility for errors or omissions. No liability is assumed for incidental or consequential damages in connection with or arising out of the use of the information or programs contained herein.

The *Unicode Character Database* and other files are provided as-is by Unicode®, Inc. No claims are made as to fitness for any particular purpose. No warranties of any kind are expressed or implied. The recipient agrees to determine applicability of information provided.

Dai Kan-Wa Jiten used as the source of reference Kanji codes was written by Tetsuji Morohashi and published by Taishukan Shoten.

ISBN 0-201-61633-5

Copyright © 1991-2000 by Unicode, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher or Unicode, Inc.

This book is set in Minion, designed by Rob Slimbach at Adobe Systems, Inc. It was typeset using FrameMaker 5.5 running under Windows NT. ASMUS, Inc. created custom software for chart layout. The Han radical-stroke index was typeset by Apple Computer, Inc. The following companies and organizations supplied fonts:

Apple Computer, Inc.
Atelier Fluxus Virus
Beijing Zhong Yi (Zheng Code) Electronics Company
DecoType, Inc.
IBM Corporation
Monotype Typography, Inc.
Microsoft Corporation
Peking University Founder Group Corporation
Production First Software

Additional fonts were supplied by individuals as listed in the *Acknowledgments*.

The Unicode® Consortium is a registered trademark, and Unicode™ is a trademark of Unicode, Inc. The Unicode logo is a trademark of Unicode, Inc., and may be registered in some jurisdictions.

All other company and product names are trademarks or registered trademarks of the company or manufacturer, respectively.

The publisher offers discounts on this book when ordered in quantity for special sales. For more information please contact:

Corporate, Government, and Special Sales
Addison Wesley Longman, Inc.
One Jacob Way
Reading, Massachusetts 01867

Visit A-W on the Web: http://www.awl.com/cseng/

First printing, January 2000.