

This PDF file is an excerpt from The Unicode Standard, Version 5.2, issued and published by the Unicode Consortium. The PDF files have not been modified to reflect the corrections found on the Updates and Errata page (<http://www.unicode.org/errata/>). For information about more recent versions of the Unicode Standard see <http://www.unicode.org/versions/enumeratedversions.html>.

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 the publisher was aware of a trademark claim, the designations have been printed with initial capital letters or in all capitals.

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.

The authors and publisher have taken care in the 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.

Copyright © 1991–2009 Unicode, Inc.

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording, or likewise. For information regarding permissions, inquire at <http://www.unicode.org/reporting.html>. For terms of use, please see <http://www.unicode.org/copyright.html>.

Visit the Unicode Consortium on the Web: <http://www.unicode.org>

The Unicode Standard / the Unicode Consortium ; edited by Julie D. Allen ... [et al.]. — Version 5.2.

Includes bibliographical references and index.

ISBN 978-1-936213-00-9 (<http://www.unicode.org/versions/Unicode5.2.0/>)

I. Unicode (Computer character set) I. Allen, Julie D.

II. Unicode Consortium.

QA268.U545 2009

ISBN 978-1-936213-00-9

Published in Mountain View, CA

December 2009

Tables

Table 2-1.	The 10 Unicode Design Principles	11
Table 2-2.	User-Perceived Characters with Multiple Code Points	13
Table 2-3.	Types of Code Points	23
Table 2-4.	The Seven Unicode Encoding Schemes	31
Table 2-5.	Interaction of Combining Characters	44
Table 2-6.	Nondefault Stacking	45
Table 3-1.	Named Unicode Algorithms	68
Table 3-2.	Normative Character Properties	73
Table 3-3.	Informative Character Properties	74
Table 3-4.	Examples of Unicode Encoding Forms	91
Table 3-5.	UTF-16 Bit Distribution	92
Table 3-6.	UTF-8 Bit Distribution	93
Table 3-7.	Well-Formed UTF-8 Byte Sequences	93
Table 3-8.	Use of U+FFFD in UTF-8 Conversion	95
Table 3-9.	Summary of UTF-16BE, UTF-16LE, and UTF-16	97
Table 3-10.	Summary of UTF-32BE, UTF-32LE, and UTF-32	98
Table 3-11.	Combining Marks and Starter Status	101
Table 3-12.	Reorderable Pairs	102
Table 3-13.	Hangul Syllable No-Break Rules	106
Table 3-14.	Korean Syllable Break Examples	107
Table 3-15.	Context Specification for Casing	111
Table 3-16.	Case Detection Examples	113
Table 4-1.	Relationship of Casing Definitions	119
Table 4-2.	Case Function Values for Strings	119
Table 4-3.	Sources for Case Mapping Information	120
Table 4-4.	Class Zero Combining Marks—Reordrant	122
Table 4-5.	Thai, Lao, and Tai Viet Logical Order Exceptions	122
Table 4-7.	Class Zero Combining Marks—Subjoined	123
Table 4-6.	Class Zero Combining Marks—Split	123
Table 4-8.	Class Zero Combining Marks—Strikethrough	124
Table 4-9.	General Category	125
Table 4-11.	Ideographs Used as Accounting Numbers	127
Table 4-10.	Primary Numeric Ideographs	127
Table 4-12.	Construction of Code Point Labels	131
Table 4-13.	Unusual Properties	133
Table 5-1.	Hex Values for Acronyms	146
Table 5-2.	NLF Platform Correlations	147
Table 5-3.	Typing Order Differing from Canonical Order	157
Table 5-4.	Permuting Combining Class Weights	158
Table 5-5.	Casing and Normalization in Strings	169
Table 5-6.	Paired Stateful Controls	174
Table 5-7.	Paired Stateful Controls (Deprecated)	175
Table 6-1.	Typology of Scripts in the Unicode Standard	181
Table 6-2.	Unicode Space Characters	184
Table 6-3.	Unicode Dash Characters	186
Table 6-4.	East Asian Quotation Marks	188
Table 6-5.	Opening and Closing Forms	189

Table 6-6.	Names for the @	193
Table 7-1.	Nonspacing Marks Used with Greek	210
Table 7-2.	Greek Spacing and Nonspacing Pairs	214
Table 8-2.	Glyph Variation in Eastern Arabic-Indic Digits	241
Table 8-1.	Arabic Digit Names	241
Table 8-3.	Primary Arabic Joining Types	243
Table 8-4.	Derived Arabic Joining Types	244
Table 8-5.	Arabic Glyph Types	244
Table 8-6.	Arabic Obligatory Ligature Joining Groups	245
Table 8-7.	Arabic Ligature Notation	246
Table 8-8.	Dual-Joining Arabic Characters	247
Table 8-9.	Right-Joining Arabic Characters	248
Table 8-10.	Forms of the Arabic Letter yeh	249
Table 8-11.	Miscellaneous Syriac Diacritic Use	254
Table 8-12.	Syriac Final Alaph Glyph Types	254
Table 8-13.	Dual-Joining Syriac Characters	255
Table 8-14.	Right-Joining Syriac Characters	256
Table 8-16.	Syriac Ligatures	257
Table 8-15.	Syriac Alaph Glyph Forms	257
Table 8-17.	Samaritan Performative Punctuation Marks	259
Table 8-18.	Thaana Glyph Placement	259
Table 9-1.	Devanagari Vowel Letters	264
Table 9-2.	Sample Devanagari Half-Forms	273
Table 9-3.	Sample Devanagari Ligatures	273
Table 9-4.	Sample Devanagari Half-Ligature Forms	274
Table 9-5.	Prishthamatra Orthography	277
Table 9-6.	Bengali Vowel Letters	279
Table 9-7.	Bengali Consonant-Vowel Combinations	280
Table 9-8.	Use of Apostrophe in Bangla	283
Table 9-9.	Gurmukhi Vowel Letters	284
Table 9-10.	Gurmukhi Conjuncts	285
Table 9-11.	Additional Pairin and Addha Forms in Gurmukhi	285
Table 9-12.	Use of Joiners in Gurmukhi	286
Table 9-14.	Gujarati Conjuncts	287
Table 9-13.	Gujarati Vowel Letters	287
Table 9-16.	Oriya Conjuncts	288
Table 9-17.	Oriya Vowel Placement	288
Table 9-15.	Oriya Vowel Letters	288
Table 9-18.	Tamil Vowel Reordering	290
Table 9-19.	Tamil Vowel Splitting and Reordering	291
Table 9-20.	Tamil Ligatures with u	292
Table 9-21.	Tamil Vowels, Consonants, and Syllables	295
Table 9-22.	Telugu Vowel Letters	296
Table 9-23.	Kannada Vowel Letters	298
Table 9-24.	Malayalam Vowel Letters	300
Table 9-25.	Malayalam Orthographic Reform	300
Table 9-26.	Malayalam Conjuncts	301
Table 9-27.	Candrakala Examples	301
Table 9-28.	Atomic Encoding of Malayalam Chillus	302
Table 9-30.	Malayalam /nr/ and /nt/	303
Table 9-29.	Malayalam /rr/ and /tt/	303
Table 10-1.	Sinhala Vowel Letters	307
Table 10-2.	Lepcha Syllabic Structure	317
Table 10-3.	Phags-pa Positional Forms of I, U, E, and O	321

Table 10-4.	Contextual Glyph Mirroring in Phags-pa	322
Table 10-5.	Phags-pa Standardized Variants	322
Table 10-6.	Positions of Limbu Combining Marks	325
Table 10-7.	Kharoshthi Vowel Signs	334
Table 10-8.	Kharoshthi Vowel Modifiers	335
Table 10-10.	Examples of Kharoshthi Virama	336
Table 10-9.	Kharoshthi Consonant Modifiers	336
Table 11-1.	Glyph Positions in Thai Syllables	339
Table 11-2.	Glyph Positions in Lao Syllables	341
Table 11-3.	Myanmar Syllabic Structure	345
Table 11-4.	Khamti Shan Tone Marks	346
Table 11-5.	Independent Khmer Vowel Characters	348
Table 11-6.	Two Registers of Khmer Consonants	350
Table 11-7.	Khmer Subscript Consonant Signs	351
Table 11-9.	Khmer Subscript Independent Vowel Signs	353
Table 11-8.	Khmer Composite Dependent Vowel Signs with Nikahit	353
Table 11-10.	Tai Le Tone Marks	357
Table 11-11.	Myanmar Digits	358
Table 11-12.	New Tai Lue Vowel Placement	359
Table 11-13.	New Tai Lue Registers and Tones	359
Table 11-14.	Tai Viet Symbols and Punctuation	363
Table 11-15.	Cham Syllabic Structure	365
Table 11-16.	Hanunóo and Buhid Vowel Sign Combinations	367
Table 11-17.	Balinese Base Consonants and Conjunct Forms	369
Table 11-18.	Sasak Extensions for Balinese	370
Table 11-19.	Balinese Consonant Clusters with u and u:	371
Table 11-20.	Sundanese Syllabic Structure	377
Table 12-1.	Sources for Unified Han	381
Table 12-2.	Blocks Containing Han Ideographs	383
Table 12-3.	Small Extensions to the URO	384
Table 12-4.	Common Han Characters	385
Table 12-5.	Source Encoding for Sword Variants	389
Table 12-6.	Ideographs Not Unified	391
Table 12-7.	Ideographs Unified	391
Table 12-8.	Han Ideograph Arrangement	392
Table 12-9.	Mandarin Tone Marks	400
Table 12-10.	Minnan and Hakka Tone Marks	401
Table 12-12.	Line-Based Placement of Jungseong	405
Table 12-11.	Separating Jamo Characters	405
Table 13-1.	Labialized Forms in Ethiopic -WAA	410
Table 13-2.	Labialized Forms in Ethiopic -WE	411
Table 13-3.	N’Ko Tone Diacritics on Vowels	422
Table 13-4.	Other N’Ko Diacritic Usage	423
Table 13-5.	N’Ko Letter Shaping	424
Table 13-6.	IPA Transcription of Deseret	431
Table 13-7.	Lisu Tone Letters	433
Table 13-8.	Punctuation Adopted in Lisu Orthography	434
Table 14-1.	Similar Characters in Linear B and Cypriot	444
Table 14-2.	Old South Arabian Numeric Characters	446
Table 14-3.	Building Numeric Values in Old South Arabian	446
Table 14-4.	Inscriptional Parthian Shaping Behavior	449
Table 14-5.	Avestan Shaping Behavior	450
Table 14-6.	Cuneiform Script Usage	454
Table 14-7.	Hieroglyphic Character Sequence	456

Table 15-1.	Currency Symbols Encoded in Other Blocks	461
Table 15-2.	Mathematical Alphanumeric Symbols	465
Table 15-3.	Use of Mathematical Symbol Pieces	478
Table 15-4.	Japanese Era Names	490
Table 15-5.	Examples of Ornamentation	495
Table 15-6.	Representation of Ancient Greek Vocal and Instrumental Notation	497
Table 16-1.	Control Codes Specified in the Unicode Standard	501
Table 16-2.	Letter Spacing	502
Table 16-3.	Bidirectional Ordering Controls	508
Table 16-4.	Unicode Encoding Scheme Signatures	516
Table 16-5.	U+FEFF Signature in Other Charsets	517
Table 17-1.	IRG Sources	531
Table A-1.	Extended BNF	539
Table A-2.	Character Class Examples	540
Table A-3.	Operators	540
Table C-1.	Timeline	550
Table C-2.	Zero Extending	553
Table D-1.	Versions of Unicode and ISO/IEC 10646-1	557
Table D-2.	Allocation of Code Points by Type	558
Table D-3.	Allocation of Code Points by Type (Early Versions)	558
Table D-4.	Version 5.1 Clause and Definition Updates	559
Table D-5.	Version 5.2 Clause and Definition Updates	559