

# The Unicode Standard

## Version 6.1 – Core Specification

To learn about the latest version of the Unicode Standard, see <http://www.unicode.org/versions/latest/>.

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.

Unicode and the Unicode Logo are registered trademarks of Unicode, Inc., in the United States and other countries.

The authors and publisher have taken care in the preparation of this specification, 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 © 1991–2012 Unicode, Inc.

All rights reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction. For information regarding permissions, inquire at <http://www.unicode.org/reporting.html>. For information about the Unicode terms of use, please see <http://www.unicode.org/copyright.html>.

The Unicode Standard / the Unicode Consortium ; edited by Julie D. Allen ... [et al.]. — Version 6.1.

Includes bibliographical references and index.

ISBN 978-1-936213-02-3 (<http://www.unicode.org/versions/Unicode6.1.0/>)

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

QA268.U545 2012

ISBN 978-1-936213-02-3

Published in Mountain View, CA

April 2012

# Tables

Table 2-1.	The 10 Unicode Design Principles . . . . .	11
Table 2-2.	User-Perceived Characters with Multiple Code Points . . . . .	12
Table 2-3.	Types of Code Points . . . . .	23
Table 2-4.	The Seven Unicode Encoding Schemes . . . . .	30
Table 2-5.	Interaction of Combining Characters . . . . .	44
Table 2-6.	Nondefault Stacking . . . . .	45
Table 3-1.	Named Unicode Algorithms . . . . .	69
Table 3-2.	Normative Character Properties . . . . .	74
Table 3-3.	Informative Character Properties . . . . .	75
Table 3-4.	Examples of Unicode Encoding Forms . . . . .	93
Table 3-5.	UTF-16 Bit Distribution . . . . .	94
Table 3-6.	UTF-8 Bit Distribution . . . . .	94
Table 3-7.	Well-Formed UTF-8 Byte Sequences . . . . .	95
Table 3-8.	Use of U+FFFD in UTF-8 Conversion . . . . .	97
Table 3-9.	Summary of UTF-16BE, UTF-16LE, and UTF-16 . . . . .	99
Table 3-10.	Summary of UTF-32BE, UTF-32LE, and UTF-32 . . . . .	100
Table 3-11.	Combining Marks and Starter Status . . . . .	103
Table 3-12.	Reorderable Pairs . . . . .	104
Table 3-13.	Hangul Characters Used in Examples . . . . .	109
Table 3-14.	Context Specification for Casing . . . . .	116
Table 3-15.	Case Detection Examples . . . . .	119
Table 4-1.	Relationship of Casing Definitions . . . . .	125
Table 4-2.	Case Function Values for Strings . . . . .	125
Table 4-3.	Sources for Case Mapping Information . . . . .	126
Table 4-4.	Class Zero Combining Marks—Reordrant . . . . .	128
Table 4-5.	Thai, Lao, and Tai Viet Logical Order Exceptions . . . . .	128
Table 4-6.	Class Zero Combining Marks—Split . . . . .	129
Table 4-7.	Class Zero Combining Marks—Subjoined . . . . .	130
Table 4-8.	Class Zero Combining Marks—Strikethrough . . . . .	130
Table 4-9.	General Category . . . . .	131
Table 4-10.	Primary Numeric Ideographs . . . . .	134
Table 4-11.	Ideographs Used as Accounting Numbers . . . . .	134
Table 4-12.	Construction of Code Point Labels . . . . .	138
Table 4-13.	Unusual Properties . . . . .	141
Table 5-1.	Hex Values for Acronyms . . . . .	154
Table 5-2.	NLF Platform Correlations . . . . .	154
Table 5-3.	Typing Order Differing from Canonical Order . . . . .	164
Table 5-4.	Permuting Combining Class Weights . . . . .	165
Table 5-5.	Casing and Normalization in Strings . . . . .	177
Table 6-1.	Typology of Scripts in the Unicode Standard . . . . .	191
Table 6-2.	Unicode Space Characters . . . . .	194
Table 6-3.	Unicode Dash Characters . . . . .	196
Table 6-4.	East Asian Quotation Marks . . . . .	199
Table 6-5.	Opening and Closing Forms . . . . .	199
Table 6-6.	Names for the @ . . . . .	203
Table 6-7.	Unicode Danda Characters . . . . .	206
Table 7-1.	Nonspacing Marks Used with Greek . . . . .	222
Table 7-2.	Greek Spacing and Nonspacing Pairs . . . . .	226

Table 8-1.	Arabic Digit Names . . . . .	254
Table 8-2.	Glyph Variation in Eastern Arabic-Indic Digits . . . . .	254
Table 8-3.	Primary Arabic Joining Types . . . . .	257
Table 8-4.	Derived Arabic Joining Types . . . . .	257
Table 8-5.	Arabic Glyph Types . . . . .	257
Table 8-6.	Arabic Obligatory Ligature Joining Groups . . . . .	259
Table 8-7.	Arabic Ligature Notation . . . . .	259
Table 8-8.	Dual-Joining Arabic Characters . . . . .	260
Table 8-9.	Right-Joining Arabic Characters . . . . .	261
Table 8-10.	Forms of the Arabic Letter yeh . . . . .	262
Table 8-11.	Arabic Letters With Hamza Above . . . . .	264
Table 8-12.	Miscellaneous Syriac Diacritic Use . . . . .	270
Table 8-13.	Syriac Final Alaph Glyph Types . . . . .	270
Table 8-14.	Dual-Joining Syriac Characters . . . . .	271
Table 8-15.	Right-Joining Syriac Characters . . . . .	272
Table 8-16.	Syriac Alaph Glyph Forms . . . . .	272
Table 8-17.	Syriac Ligatures . . . . .	272
Table 8-18.	Samaritan Performative Punctuation Marks . . . . .	274
Table 8-19.	Thaana Glyph Placement . . . . .	275
Table 9-1.	Devanagari Vowel Letters . . . . .	281
Table 9-2.	Sample Devanagari Half-Forms . . . . .	289
Table 9-3.	Sample Devanagari Ligatures . . . . .	290
Table 9-4.	Sample Devanagari Half-Ligature Forms . . . . .	291
Table 9-5.	Devanagari Vowels Used in Bihari Languages . . . . .	293
Table 9-6.	Prishthamatra Orthography . . . . .	293
Table 9-7.	Bengali Vowel Letters . . . . .	296
Table 9-8.	Bengali Consonant-Vowel Combinations . . . . .	297
Table 9-9.	Use of Apostrophe in Bangla . . . . .	299
Table 9-10.	Gurmukhi Vowel Letters . . . . .	301
Table 9-11.	Gurmukhi Conjuncts . . . . .	302
Table 9-12.	Additional Pairin and Addha Forms in Gurmukhi . . . . .	302
Table 9-13.	Use of Joiners in Gurmukhi . . . . .	303
Table 9-14.	Gujarati Vowel Letters . . . . .	303
Table 9-15.	Gujarati Conjuncts . . . . .	304
Table 9-16.	Oriya Vowel Letters . . . . .	304
Table 9-17.	Oriya Conjuncts . . . . .	305
Table 9-18.	Oriya Vowel Placement . . . . .	305
Table 9-19.	Tamil Vowel Reordering . . . . .	307
Table 9-20.	Tamil Vowel Splitting and Reordering . . . . .	308
Table 9-21.	Tamil Ligatures with u . . . . .	309
Table 9-22.	Tamil Vowels, Consonants, and Syllables . . . . .	312
Table 9-23.	Telugu Vowel Letters . . . . .	313
Table 9-24.	Rendering of Telugu na + virama . . . . .	314
Table 9-25.	Kannada Vowel Letters . . . . .	316
Table 9-26.	Malayalam Vowel Letters . . . . .	318
Table 9-27.	Malayalam Orthographic Reform . . . . .	318
Table 9-28.	Malayalam Conjuncts . . . . .	319
Table 9-29.	Candrakala Examples . . . . .	319
Table 9-30.	Atomic Encoding of Malayalam Chillus . . . . .	320
Table 9-31.	Malayalam /rr/ and /tt/ . . . . .	320
Table 9-32.	Malayalam /nr/ and /nt/ . . . . .	321
Table 10-1.	Sinhala Vowel Letters . . . . .	325
Table 10-2.	Lepcha Syllabic Structure . . . . .	336
Table 10-3.	Phags-pa Positional Forms of I, U, E, and O . . . . .	340

Table 10-4.	Contextual Glyph Mirroring in Phags-pa . . . . .	340
Table 10-5.	Phags-pa Standardized Variants . . . . .	341
Table 10-6.	Positions of Limbu Combining Characters . . . . .	344
Table 10-7.	Takri Vowel Letters . . . . .	350
Table 10-8.	Kharoshthi Vowel Signs . . . . .	357
Table 10-9.	Kharoshthi Vowel Modifiers . . . . .	358
Table 10-10.	Kharoshthi Consonant Modifiers . . . . .	358
Table 10-11.	Examples of Kharoshthi Virama . . . . .	359
Table 10-12.	Brahmi Vowel Letters . . . . .	359
Table 10-13.	Brahmi Positional Digits . . . . .	361
Table 11-1.	Glyph Positions in Thai Syllables . . . . .	365
Table 11-2.	Glyph Positions in Lao Syllables . . . . .	367
Table 11-3.	Myanmar Syllabic Structure . . . . .	371
Table 11-4.	Khamti Shan Tone Marks . . . . .	372
Table 11-5.	Independent Khmer Vowel Characters . . . . .	374
Table 11-6.	Two Registers of Khmer Consonants . . . . .	376
Table 11-7.	Khmer Subscript Consonant Signs . . . . .	377
Table 11-8.	Khmer Composite Dependent Vowel Signs with Nikahit . . . . .	378
Table 11-9.	Khmer Subscript Independent Vowel Signs . . . . .	379
Table 11-10.	Tai Le Tone Marks . . . . .	383
Table 11-11.	Myanmar Digits . . . . .	384
Table 11-12.	New Tai Lue Vowel Placement . . . . .	385
Table 11-13.	New Tai Lue Registers and Tones . . . . .	385
Table 11-14.	Tai Viet Symbols and Punctuation . . . . .	389
Table 11-15.	Cham Syllabic Structure . . . . .	391
Table 11-16.	Hanunóo and Buhid Vowel Sign Combinations . . . . .	393
Table 11-17.	Balinese Base Consonants and Conjunct Forms . . . . .	395
Table 11-18.	Sasak Extensions for Balinese . . . . .	396
Table 11-19.	Balinese Consonant Clusters with u and u: . . . . .	397
Table 11-20.	Sundanese Syllabic Structure . . . . .	404
Table 12-1.	Sources for Unified Han . . . . .	407
Table 12-2.	Blocks Containing Han Ideographs . . . . .	409
Table 12-3.	Small Extensions to the URO . . . . .	410
Table 12-4.	Common Han Characters . . . . .	411
Table 12-5.	Source Encoding for Sword Variants . . . . .	416
Table 12-6.	Ideographs Not Unified . . . . .	417
Table 12-7.	Ideographs Unified . . . . .	418
Table 12-8.	Han Ideograph Arrangement . . . . .	418
Table 12-9.	Mandarin Tone Marks . . . . .	426
Table 12-10.	Minnan and Hakka Tone Marks . . . . .	427
Table 12-11.	Separating Jamo Characters . . . . .	431
Table 12-12.	Line-Based Placement of Jungseong . . . . .	432
Table 13-1.	Labialized Forms in Ethiopic -WAA . . . . .	438
Table 13-2.	Labialized Forms in Ethiopic -WE . . . . .	439
Table 13-3.	N’Ko Tone Diacritics on Vowels . . . . .	451
Table 13-4.	Other N’Ko Diacritic Usage . . . . .	452
Table 13-5.	N’Ko Letter Shaping . . . . .	453
Table 13-6.	IPA Transcription of Deseret . . . . .	460
Table 13-7.	Lisu Tone Letters . . . . .	462
Table 13-8.	Punctuation Adopted in Lisu Orthography . . . . .	463
Table 14-1.	Similar Characters in Linear B and Cypriot . . . . .	474
Table 14-2.	Old South Arabian Numeric Characters . . . . .	476
Table 14-3.	Number Formation in Old South Arabian . . . . .	477
Table 14-4.	Number Formation in Aramaic . . . . .	479

Table 14-5.	Dual-Joining Mandaic Characters . . . . .	480
Table 14-6.	Right-Joining Mandaic Characters . . . . .	481
Table 14-7.	Inscriptional Parthian Shaping Behavior . . . . .	482
Table 14-8.	Avestan Shaping Behavior . . . . .	483
Table 14-9.	Cuneiform Script Usage . . . . .	486
Table 14-10.	Hieroglyphic Character Sequence . . . . .	489
Table 15-1.	Currency Symbols Encoded in Other Blocks . . . . .	495
Table 15-2.	Mathematical Alphanumeric Symbols . . . . .	499
Table 15-3.	Script-Specific Decimal Digits . . . . .	502
Table 15-4.	Compatibility Digits . . . . .	504
Table 15-5.	Use of Mathematical Symbol Pieces . . . . .	519
Table 15-6.	Japanese Era Names . . . . .	533
Table 15-7.	Examples of Ornamentation . . . . .	539
Table 15-8.	Representation of Ancient Greek Vocal and Instrumental Notation	541
Table 16-1.	Control Codes Specified in the Unicode Standard . . . . .	545
Table 16-2.	Letter Spacing . . . . .	547
Table 16-3.	Bidirectional Ordering Controls . . . . .	553
Table 16-4.	Paired Stateful Controls . . . . .	554
Table 16-5.	Paired Stateful Controls (Deprecated) . . . . .	554
Table 16-6.	Unicode Encoding Scheme Signatures . . . . .	562
Table 16-7.	U+FEFF Signature in Other Charsets . . . . .	563
Table 17-1.	IRG Sources . . . . .	577
Table A-1.	Extended BNF . . . . .	583
Table A-2.	Character Class Examples . . . . .	584
Table A-3.	Operators . . . . .	584
Table C-1.	Timeline . . . . .	594
Table C-2.	Zero Extending . . . . .	598
Table D-1.	Versions of Unicode and ISO/IEC 10646-1 . . . . .	601
Table D-2.	Allocation of Code Points by Type . . . . .	602
Table D-3.	Allocation of Code Points by Type (Early Versions) . . . . .	602
Table D-4.	Version 5.1 Clause and Definition Updates . . . . .	603
Table D-5.	Version 5.2 Clause and Definition Updates . . . . .	603
Table D-6.	Version 6.0 Clause and Definition Updates . . . . .	604
Table D-7.	Version 6.1 Clause and Definition Updates . . . . .	604
Table F-1.	CJK Strokes . . . . .	610