The Unicode® Standard
Version 9.0 – 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.

© 2016 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 the Unicode Consortium. — Version 9.0.
Includes bibliographical references and index.
ISBN 978-1-936213-13-9 (http://www.unicode.org/versions/Unicode9.0.0/)
1. Unicode (Computer character set) I. Unicode Consortium.
QA268.U545 2016

ISBN 978-1-936213-13-9
Published in Mountain View, CA
July 2016

Tables

Table 2-1. The 10 Unicode Design Principles 14
Table 2-2. User-Perceived Characters with Multiple Code Points 16
Table 2-3. Types of Code Points 30
Table 2-4. The Seven Unicode Encoding Schemes 41
Table 2-5. Interaction of Combining Characters 58
Table 2-6. Nondefault Stacking 59
Table 3-1. Named Unicode Algorithms 93
Table 3-2. Normative Character Properties 99
Table 3-3. Informative Character Properties 100
Table 3-4. Examples of Unicode Encoding Forms 123
Table 3-5. UTF-16 Bit Distribution 124
Table 3-6. UTF-8 Bit Distribution 125
Table 3-7. Well-Formed UTF-8 Byte Sequences 125
Table 3-8. Use of U+FFFD in UTF-8 Conversion 128
Table 3-9. Summary of UTF-16BE, UTF-16LE, and UTF-16 131
Table 3-10. Summary of UTF-32BE, UTF-32LE, and UTF-32 132
Table 3-11. Combining Marks and Starter Status 137
Table 3-12. Reorderable Pairs 138
Table 3-13. Hangul Characters Used in Examples 144
Table 3-14. Context Specification for Casing 153
Table 3-15. Case Detection Examples 157
Table 4-1. Relationship of Casing Definitions 165
Table 4-2. Case Function Values for Strings 166
Table 4-3. Sources for Case Mapping Information 166
Table 4-4. Class Zero Combining Marks—Reordrant 169
Table 4-5. Thai, Lao, and Other Logical Order Exceptions 170
Table 4-6. Class Zero Combining Marks—Split 171
Table 4-7. Class Zero Combining Marks—Subjoined 172
Table 4-8. Class Zero Combining Marks—Strikethrough 172
Table 4-9. General Category 175
Table 4-10. Primary Numeric Ideographs 178
Table 4-11. Ideographs Used as Accounting Numbers 178
Table 4-12. Types of Character Name Aliases 183
Table 4-13. Name Derivation Rule Prefix Strings 186
Table 4-14. Construction of Code Point Labels 188
Table 4-15. Unusual Properties 192
Table 5-1. Hex Values for Acronyms 211
Table 5-2. NLF Platform Correlations 212
Table 5-3. Typing Order Differing from Canonical Order 227
Table 5-4. Permuting Combining Class Weights 227

Table 5-5. Casing and Normalization in Strings 244
Table 6-1. Typology of Scripts in the Unicode Standard 264
Table 6-2. Unicode Space Characters 268
Table 6-3. Unicode Dash Characters 270
Table 6-4. Models of Visual Relationship between Quote Glyphs 273
Table 6-5. East Asian Quotation Marks 274
Table 6-6. Opening and Closing Forms 275
Table 6-7. Names for the @ 280
Table 6-8. Unicode Danda Characters 284
Table 7-1. Preferred Rendering of Cedilla versus Comma Below 294
Table 7-2. Nonspacing Marks Used with Greek 306
Table 7-3. Greek Spacing and Nonspacing Pairs 311
Table 8-1. Similar Characters in Linear B and Cypriot 344
Table 8-2. Combining Marks Used in Old Permic 357
Table 9-1. Arabic Digit Names 373
Table 9-2. Glyph Variation in Eastern Arabic-Indic Digits 374
Table 9-3. Primary Arabic Joining Types 377
Table 9-4. Derived Arabic Joining Types 378
Table 9-5. Arabic Glyph Types 379
Table 9-6. Arabic Obligatory Ligature Joining Groups 380
Table 9-7. Arabic Ligature Notation 381
Table 9-8. Dual-Joining Arabic Characters 382
Table 9-9. Right-Joining Arabic Characters 384
Table 9-10. Forms of the Arabic Letter yeh 385
Table 9-11. Arabic Letters With Hamza Above 388
Table 9-12. Miscellaneous Syriac Diacritic Use 396
Table 9-13. Syriac Final Alaph Glyph Types 397
Table 9-14. Dual-Joining Syriac Characters 398
Table 9-15. Right-Joining Syriac Characters 399
Table 9-16. Syriac Alaph Glyph Forms 399
Table 9-17. Syriac Ligatures 400
Table 9-18. Samaritan Performative Punctuation Marks 402
Table 9-19. Dual-Joining Mandaic Characters 404
Table 9-20. Right-Joining Mandaic Characters 405
Table 10-1. Old South Arabian Numeric Characters 411
Table 10-2. Number Formation in Old South Arabian 411
Table 10-3. Number Formation in Aramaic 414
Table 10-4. Dual-Joining Manichaean Letters 417
Table 10-5. Right-Joining Manichaean Letters 417
Table 10-6. Left-Joining Manichaean Letters 418
Table 10-7. Non-Joining Manichaean Letters 418
Table 10-8. Manichaean Ligatures 418
Table 10-9. Inscriptional Parthian Shaping Behavior 421
Table 10-10. Avestan Shaping Behavior 422
Table 11-1. Cuneiform Script Usage 429

Table 11-2. Hieroglyphic Character Sequence 435
Table 12-1. Devanagari Vowel Letters 448
Table 12-2. Sample Devanagari Half-Forms 458
Table 12-3. Sample Devanagari Ligatures 459
Table 12-4. RA + Vocalic Letter Ligature Forms 460
Table 12-5. Sample Devanagari Half-Ligature Forms 460
Table 12-6. Marathi and Nepali Allographs 461
Table 12-7. Devanagari Vowels Used in Bihari Languages 463
Table 12-8. Prishthamatra Orthography 464
Table 12-9. Bengali Vowel Letters 467
Table 12-10. Diphthong Vowel Letters in Kokborok 468
Table 12-11. Assamese Consonant-Vowel Combinations 468
Table 12-12. Bengali Consonant-Vowel Combinations 469
Table 12-13. Use of Apostrophe in Bangla 472
Table 12-14. Gurmukhi Vowel Letters 474
Table 12-15. Gurmukhi Conjuncts 475
Table 12-16. Additional Pairin and Addha Forms in Gurmukhi 476
Table 12-17. Use of Joiners in Gurmukhi 476
Table 12-18. Gujarati Vowel Letters 478
Table 12-19. Gujarati Conjuncts 479
Table 12-20. Oriya Vowel Letters 480
Table 12-21. Oriya Conjuncts 481
Table 12-22. Oriya Vowel Placement 481
Table 12-23. Ligation for the Syllable om 482
Table 12-24. Tamil Ligatures with u 486
Table 12-25. Tamil Vowels, Consonants, and Syllables 490
Table 12-26. Telugu Vowel Letters 492
Table 12-27. Rendering of Telugu na + virama 493
Table 12-28. Kannada Vowel Letters 495
Table 12-29. Rendering of Kannada na + virama 498
Table 12-30. Malayalam Vowel Letters 499
Table 12-31. Malayalam Orthographic Reform 500
Table 12-32. Malayalam Conjuncts 501
Table 12-33. Candrakkala Examples 501
Table 12-34. Use of Joiners in Malayalam 502
Table 12-35. Malayalam /ra̲ra̲/ and /t̲t̲a/ 504
Table 12-36. Malayalam /nr/ and /nt/ 504
Table 12-37. Atomic Encoding of Malayalam Chillus 505
Table 13-1. Thaana Glyph Placement 509
Table 13-2. Sinhala Vowel Letters 511
Table 13-3. Murmured Resonants in Nepal Bhasa 514
Table 13-4. Positions of Limbu Combining Characters 538
Table 13-5. Lepcha Syllabic Structure 548
Table 14-1. Brahmi Vowel Letters 552
Table 14-2. Brahmi Positional Digits 555

Table 14-3. Kharoshthi Vowel Signs 558
Table 14-4. Kharoshthi Vowel Modifiers 560
Table 14-5. Kharoshthi Consonant Modifiers 560
Table 14-6. Examples of Kharoshthi Virama 561
Table 14-7. Phags-pa Positional Forms of I, U, E, and O 567
Table 14-8. Contextual Glyph Mirroring in Phags-pa 568
Table 14-9. Phags-pa Standardized Variants 569
Table 15-1. Takri Vowel Letters 583
Table 15-2. Siddham Punctuation Characters 585
Table 15-3. Khudawadi Vowel Letters 590
Table 15-4. Representation of Arabic Sounds in Khudawadi 591
Table 15-5. Tirhuta Vowel Letters 594
Table 15-6. Modi Vowel Letters 597
Table 15-7. Rendering of Explicit Virama Forms in Grantha 600
Table 15-8. Additional Svara Marks used in Grantha 601
Table 16-1. Glyph Positions in Thai Syllables 608
Table 16-2. Glyph Positions in Lao Syllables 611
Table 16-3. Modern Burmese Syllabic Structure 616
Table 16-4. Khamti Shan Tone Marks 619
Table 16-5. Independent Khmer Vowel Characters 622
Table 16-6. Two Registers of Khmer Consonants 623
Table 16-7. Khmer Subscript Consonant Signs 624
Table 16-8. Khmer Composite Dependent Vowel Signs with Nikahit 626
Table 16-9. Khmer Subscript Independent Vowel Signs 627
Table 16-10. Tai Le Tone Marks 632
Table 16-11. Myanmar Digits in Tai Le 633
Table 16-12. New Tai Lue Vowel Placement 635
Table 16-13. New Tai Lue Registers and Tones 636
Table 16-14. Tai Viet Symbols and Punctuation 641
Table 16-15. Cham Syllabic Structure 645
Table 17-1. Hanunóo and Buhid Vowel Sign Combinations 651
Table 17-2. Balinese Base Consonants and Conjunct Forms 655
Table 17-3. Sasak Extensions for Balinese 657
Table 17-4. Balinese Consonant Clusters with u and u: 659
Table 17-5. Modern Sundanese Syllabic Structure 668
Table 18-1. Blocks Containing Han Ideographs 672
Table 18-2. Small Extensions to the URO 672
Table 18-3. Common Han Characters 674
Table 18-4. Source Encoding for Sword Variants 680
Table 18-5. Ideographs Not Unified 682
Table 18-6. Ideographs Unified 682
Table 18-7. Han Ideograph Arrangement 683
Table 18-8. Mandarin Tone Marks 693
Table 18-9. Minnan and Hakka Tone Marks 694
Table 18-10. Separating Jamo Characters 700

Table 18-11. Line-Based Placement of Jungseong 702
Table 18-12. Lisu Tone Letters 707
Table 18-13. Punctuation Adopted in Lisu Orthography 708
Table 19-1. Labialized Forms in Ethiopic -WAA 715
Table 19-2. Labialized Forms in Ethiopic -WE 715
Table 19-3. N'Ko Diacritic Usage 723
Table 19-4. N'Ko Tone Diacritics on Vowels 724
Table 19-5. N'Ko Letter Shaping 725
Table 19-6. Number Formation in Mende Kikakui 732
Table 20-1. Combining Marks used in Osage 742
Table 20-2. IPA Transcription of Deseret 745
Table 21-1. Examples of Ornamentation 754
Table 21-2. Representation of Ancient Greek Vocal and Instrumental Notation 756
Table 22-1. Currency Symbols Encoded in Other Blocks 766
Table 22-2. Mathematical Alphanumeric Symbols 772
Table 22-3. Script-Specific Decimal Digits 776
Table 22-4. Compatibility Digits 778
Table 22-5. Mathematical Operators Disunified from Punctuation 790
Table 22-6. Use of Mathematical Symbol Pieces 799
Table 22-7. Geometric Shape Collections 804
Table 22-8. Japanese Era Names 819
Table 23-1. Control Codes Specified in the Unicode Standard 823
Table 23-2. Letter Spacing 826
Table 23-3. Bidirectional Ordering Controls 833
Table 23-4. Paired Stateful Controls 834
Table 23-5. Paired Stateful Controls (Deprecated) 835
Table 23-6. Unicode Encoding Scheme Signatures 848
Table 23-7. U+FEFF Signature in Other Charsets 849
Table 24-1. IRG Sources 870
Table A-1. Extended BNF 877
Table A-2. Character Class Examples 879
Table A-3. Operators 879
Table C-1. Timeline 895
Table C-2. Zero Extending 900
Table D-1. Versions of Unicode and ISO/IEC 10646 907
Table D-2. Allocation of Code Points by Type (Versions 1.0.0 to 3.0) 908
Table D-3. Allocation of Code Points by Type (Versions 3.1 to 5.1) 909
Table D-4. Allocation of Code Points by Type (Versions 5.2 to 7.0) 910
Table D-5. Allocation of Code Points by Type (Versions 8.0 to 9.0) 911
Table E-1. G Source Documentation 919
Table E-2. H Source Documentation 920
Table E-3. M Source Documentation 920
Table E-4. T Source Documentation 920
Table E-5. J Source Documentation 921
Table E-6. K Source Documentation 921

Table E-7. KP Source Documentation 921
Table E-8. V Source Documentation 922
Table E-9. U Source Documentation 922
Table F-1. CJK Strokes 924