

# The Unicode Standard

## Version 8.0 – 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.

Copyright © 1991–2015 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 8.0

Includes bibliographical references and index.

ISBN 978-1-936213-10-8 (<http://www.unicode.org/versions/Unicode8.0.0/>)

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

QA268.U545 2015

ISBN 978-1-936213-10-8

Published in Mountain View, CA

August 2015

# Figures

Figure 1-1.	Wide ASCII. . . . .	2
Figure 1-2.	Unicode Compared to the 2022 Framework. . . . .	5
Figure 2-1.	Text Elements and Characters . . . . .	11
Figure 2-2.	Characters Versus Glyphs . . . . .	16
Figure 2-3.	Unicode Character Code to Rendered Glyphs . . . . .	17
Figure 2-4.	Bidirectional Ordering . . . . .	20
Figure 2-5.	Writing Direction and Numbers . . . . .	20
Figure 2-6.	Typeface Variation for the Bone Character. . . . .	22
Figure 2-7.	Dynamic Composition . . . . .	23
Figure 2-8.	Abstract and Encoded Characters . . . . .	29
Figure 2-9.	Overlap in Legacy Mixed-Width Encodings . . . . .	33
Figure 2-10.	Boundaries and Interpretation . . . . .	34
Figure 2-11.	Unicode Encoding Forms . . . . .	35
Figure 2-12.	Unicode Encoding Schemes . . . . .	41
Figure 2-13.	Unicode Allocation . . . . .	48
Figure 2-14.	Allocation on the BMP . . . . .	49
Figure 2-15.	Allocation on Plane 1. . . . .	51
Figure 2-16.	Writing Directions. . . . .	53
Figure 2-17.	Combining Enclosing Marks for Symbols. . . . .	56
Figure 2-18.	Sequence of Base Characters and Diacritics . . . . .	56
Figure 2-19.	Reordered Indic Vowel Signs . . . . .	57
Figure 2-20.	Properties and Combining Character Sequences . . . . .	57
Figure 2-21.	Stacking Sequences . . . . .	57
Figure 2-22.	Ligated Multiple Base Characters. . . . .	60
Figure 2-23.	Equivalent Sequences . . . . .	62
Figure 2-24.	Canonical Ordering. . . . .	63
Figure 2-25.	Types of Decomposables. . . . .	64
Figure 3-1.	Enclosing Marks. . . . .	112
Figure 4-1.	Positions of Common Combining Marks . . . . .	168
Figure 5-1.	Two-Stage Tables. . . . .	197
Figure 5-2.	Normalization . . . . .	206
Figure 5-3.	Consistent Character Boundaries. . . . .	217
Figure 5-4.	Dead Keys Versus Handwriting Sequence. . . . .	220
Figure 5-5.	Truncating Grapheme Clusters . . . . .	221
Figure 5-6.	Inside-Out Rule . . . . .	222
Figure 5-7.	Fallback Rendering . . . . .	223
Figure 5-8.	Bidirectional Placement . . . . .	224
Figure 5-9.	Justification. . . . .	224
Figure 5-10.	Positioning with Ligatures . . . . .	226
Figure 5-11.	Positioning with Contextual Forms . . . . .	227

Figure 5-12.	Positioning with Enhanced Kerning . . . . .	227
Figure 5-13.	Sublinear Searching . . . . .	232
Figure 5-14.	Uppercase Mapping for Turkish I . . . . .	238
Figure 5-15.	Lowercase Mapping for Turkish I . . . . .	238
Figure 5-16.	Casing of German Sharp S . . . . .	239
Figure 6-1.	Overriding Inherent Vowels . . . . .	260
Figure 6-2.	Forms of CJK Punctuation . . . . .	264
Figure 6-3.	European Quotation Marks . . . . .	271
Figure 6-4.	Asian Quotation Marks . . . . .	273
Figure 6-5.	Examples of Ancient Greek Editorial Marks . . . . .	281
Figure 6-6.	Use of Greek Paragraphos . . . . .	281
Figure 6-7.	CJK Parentheses . . . . .	284
Figure 7-1.	Alternative Glyphs in Latin . . . . .	291
Figure 7-2.	Diacritics on <i>i</i> and <i>j</i> . . . . .	294
Figure 7-3.	Vietnamese Letters and Tone Marks . . . . .	294
Figure 7-4.	Variations in Greek Capital Letter Upsilon . . . . .	306
Figure 7-5.	Coptic Numerals . . . . .	313
Figure 7-6.	Combination of Titlo Letters . . . . .	316
Figure 7-7.	Georgian Scripts and Casing . . . . .	321
Figure 7-8.	Tone Letters . . . . .	326
Figure 7-9.	Double Diacritics . . . . .	330
Figure 7-10.	Positioning of Double Diacritics . . . . .	330
Figure 7-11.	Use of CGJ with Double Diacritics . . . . .	330
Figure 7-12.	Interaction of Combining Marks with Ligatures . . . . .	332
Figure 7-13.	Positioning of Combining Parentheses . . . . .	333
Figure 7-14.	Use of Vertical Line Overlay for Negation . . . . .	334
Figure 7-15.	Double Diacritics and Half Marks . . . . .	335
Figure 8-1.	Distribution of Old Italic . . . . .	347
Figure 9-1.	Directionality and Cursive Connection . . . . .	367
Figure 9-2.	Using a Joiner . . . . .	369
Figure 9-3.	Using a Non-joiner . . . . .	369
Figure 9-4.	Combinations of Joiners and Non-joiners . . . . .	370
Figure 9-5.	Placement of Harakat . . . . .	370
Figure 9-6.	Arabic Year Sign . . . . .	374
Figure 9-7.	Syriac Abbreviation . . . . .	391
Figure 9-8.	Use of SAM . . . . .	391
Figure 11-1.	Interpretation of Hieroglyphic Markup . . . . .	432
Figure 12-1.	Dead Consonants in Devanagari . . . . .	445
Figure 12-2.	Conjunct Formations in Devanagari . . . . .	445
Figure 12-3.	Preventing Conjunct Forms in Devanagari . . . . .	446
Figure 12-4.	Half-Consonants in Devanagari . . . . .	447
Figure 12-5.	Independent Half-Forms in Devanagari . . . . .	447
Figure 12-6.	Half-Consonants in Oriya . . . . .	447
Figure 12-7.	Consonant Forms in Devanagari and Oriya . . . . .	448
Figure 12-8.	Rendering Order in Devanagari . . . . .	453

Figure 12-9.	Use of Apostrophe in Bodo, Dogri and Maithili. . . . .	458
Figure 12-10.	Use of Avagraha in Dogri . . . . .	459
Figure 12-11.	Requesting Bengali Consonant-Vowel Ligature . . . . .	466
Figure 12-12.	Blocking Bengali Consonant-Vowel Ligature . . . . .	466
Figure 12-13.	Bengali Syllable tta . . . . .	467
Figure 12-14.	Kssa Ligature in Tamil. . . . .	479
Figure 12-15.	Tamil Two-Part Vowels . . . . .	480
Figure 12-16.	Vowel Reordering Around a Tamil Conjunct. . . . .	481
Figure 12-17.	Tamil Ligatures with i . . . . .	482
Figure 12-18.	Spacing Forms of Tamil u . . . . .	483
Figure 12-19.	Tamil Ligatures with ra . . . . .	483
Figure 12-20.	Traditional Tamil Ligatures with aa . . . . .	483
Figure 12-21.	Traditional Tamil Ligatures with o . . . . .	484
Figure 12-22.	Traditional Tamil Ligatures with ai . . . . .	484
Figure 12-23.	Vowel ai in Modern Tamil . . . . .	484
Figure 12-24.	Indicating Retroflexion in Badaga Vowels. . . . .	493
Figure 13-1.	Tibetan Syllable Structure. . . . .	510
Figure 13-2.	Justifying Tibetan Tseks . . . . .	519
Figure 13-3.	Mongolian Glyph Convergence . . . . .	523
Figure 13-4.	Mongolian Consonant Ligation . . . . .	524
Figure 13-5.	Mongolian Positional Forms . . . . .	524
Figure 13-6.	Mongolian Free Variation Selector . . . . .	525
Figure 13-7.	Mongolian Gender Forms. . . . .	527
Figure 13-8.	Mongolian Vowel Separator . . . . .	528
Figure 14-1.	Consonant Ligatures in Brahmi . . . . .	547
Figure 14-2.	Geographical Extent of the Kharoshthi Script . . . . .	550
Figure 14-3.	Kharoshthi Number 1996 . . . . .	551
Figure 14-4.	Kharoshthi Rendering Example . . . . .	552
Figure 14-5.	Phags-pa Syllable Om . . . . .	558
Figure 14-6.	Phags-pa Reversed Shaping. . . . .	561
Figure 15-1.	Siddham Consonant Cluster. . . . .	576
Figure 15-2.	Modi Shaping for ra. . . . .	589
Figure 15-3.	Splitting Large Conjunct Stacks in Grantha . . . . .	592
Figure 16-1.	Common Ligatures in Khmer. . . . .	620
Figure 16-2.	Common Multiple Forms in Khmer . . . . .	620
Figure 16-3.	Examples of Syllabic Order in Khmer . . . . .	622
Figure 16-4.	Ligation in <i>Muul</i> Style in Khmer . . . . .	623
Figure 16-5.	Pahawh Hmong Syllable Structure. . . . .	638
Figure 17-1.	Buginese Ligature. . . . .	645
Figure 17-2.	Writing dharma in Balinese . . . . .	650
Figure 17-3.	Representation of Javanese Two-Part Vowels. . . . .	654
Figure 18-1.	Han Spelling . . . . .	668
Figure 18-2.	Semantic Context for Han Characters. . . . .	668
Figure 18-3.	Three-Dimensional Conceptual Model. . . . .	670
Figure 18-4.	CJK Source Separation . . . . .	671

Figure 18-5.	Not Cognates, Not Unified . . . . .	672
Figure 18-6.	Ideographic Component Structure . . . . .	673
Figure 18-7.	The Most Superior Node of an Ideographic Component . . . . .	673
Figure 18-8.	Using the Ideographic Description Characters . . . . .	683
Figure 18-9.	Japanese Historic Kana for e and ye . . . . .	689
Figure 19-1.	Tifinagh Contextual Shaping . . . . .	710
Figure 19-2.	Tifinagh Consonant Joiner and Bi-consonants . . . . .	711
Figure 19-3.	Examples of N’Ko Ordinals . . . . .	714
Figure 20-1.	Short Words Equivalent to Deseret Letter Names . . . . .	731
Figure 21-1.	Examples of Specialized Music Layout . . . . .	738
Figure 21-2.	Precomposed Note Characters . . . . .	739
Figure 21-3.	Alternative Noteheads . . . . .	739
Figure 21-4.	Augmentation Dots and Articulation Symbols . . . . .	739
Figure 22-1.	Alternative Glyphs for Dollar Sign . . . . .	751
Figure 22-2.	Alternative Glyphs for Numero Sign . . . . .	754
Figure 22-3.	Wide Mathematical Accents . . . . .	757
Figure 22-4.	Style Variants and Semantic Distinctions in Mathematics . . . . .	757
Figure 22-5.	Easily Confused Shapes for Mathematical Glyphs . . . . .	759
Figure 22-6.	CJK Ideographic Numbers . . . . .	763
Figure 22-7.	Regular and Old Style Digits . . . . .	765
Figure 22-8.	Alternate Forms of Vulgar Fractions . . . . .	770
Figure 22-9.	Usage of Crops and Quine Corners . . . . .	784
Figure 22-10.	Usage of the Decimal Exponent Symbol . . . . .	786
Figure 23-1.	Prevention of Joining . . . . .	815
Figure 23-2.	Exhibition of Joining Glyphs in Isolation . . . . .	815
Figure 23-3.	Effect of Intervening Joiners . . . . .	816
Figure 23-4.	Annotation Characters . . . . .	836
Figure 23-5.	Tag Characters . . . . .	840
Figure 24-1.	CJK Chart Format for the Main CJK Block . . . . .	856
Figure 24-2.	CJK Chart Format for CJK Extension A . . . . .	856
Figure 24-3.	CJK Chart Format for CJK Extension B . . . . .	856
Figure 24-4.	CJK Chart Format for Compatibility Ideographs . . . . .	857
Figure A-1.	Example of Rendering . . . . .	860