Appendix D

Changes from Unicode Version 2.0

D.1 Versions of the Unicode Standard

The Unicode Technical Committee periodically updates the Unicode Standard to respond to industry needs and to maintain consistency with ISO/IEC 10646. There have been five formally numbered major and minor versions of the Unicode Standard. The relationship between these versions of Unicode and ISO/IEC 10646 is shown in Table D-1. For more detail on the relationship of Unicode and ISO/IEC 10646, see Appendix C, Relationship to ISO/IEC 10646.

<table>
<thead>
<tr>
<th>Year</th>
<th>Version</th>
<th>Published</th>
<th>ISO/IEC 10646-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991</td>
<td>Unicode 1.0</td>
<td>Vol. 1, Addison-Wesley</td>
<td>Basis of Draft-2 ISO 10646-1</td>
</tr>
<tr>
<td>1992</td>
<td>Unicode 1.0.1</td>
<td>Vol. 1, 2, Addison-Wesley</td>
<td>Interim merger version</td>
</tr>
<tr>
<td>1993</td>
<td>Unicode 1.1</td>
<td>Technical Report #4</td>
<td>Matches ISO 10646-1</td>
</tr>
<tr>
<td>1996</td>
<td>Unicode 2.0</td>
<td>Addison-Wesley</td>
<td>Matches ISO 10646-1 plus amendments</td>
</tr>
<tr>
<td>1998</td>
<td>Unicode 2.1</td>
<td>Technical Report #8</td>
<td>Matches ISO 10646-1 plus amendments</td>
</tr>
<tr>
<td>2000</td>
<td>Unicode 3.0</td>
<td>Addison-Wesley</td>
<td>Matches ISO 10646-1 second edition</td>
</tr>
</tbody>
</table>

The Unicode Standard has grown from having 28,302 assigned character values in Version 1.0, to having 49,194 assigned character values in Version 3.0. Table D-2 documents the number of characters assigned in the different versions of the Unicode Standard.

<table>
<thead>
<tr>
<th></th>
<th>V 1.0</th>
<th>V 1.1</th>
<th>V 2.0</th>
<th>V 2.1</th>
<th>V 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alphabetics, Symbols</td>
<td>4,748</td>
<td>6,309</td>
<td>6,509</td>
<td>6,511</td>
<td>10,236</td>
</tr>
<tr>
<td>Han (URO)</td>
<td>20,902</td>
<td>20,902</td>
<td>20,902</td>
<td>20,902</td>
<td>20,902</td>
</tr>
<tr>
<td>Han Extension A</td>
<td>6,582</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Han Compatibility</td>
<td>302</td>
<td>302</td>
<td>302</td>
<td>302</td>
<td>302</td>
</tr>
<tr>
<td>Hangul Syllables</td>
<td>2,350</td>
<td>6,656</td>
<td>11,172</td>
<td>11,172</td>
<td>11,172</td>
</tr>
<tr>
<td><strong>Total assigned characters</strong></td>
<td><strong>28,302</strong></td>
<td><strong>34,169</strong></td>
<td><strong>38,885</strong></td>
<td><strong>38,887</strong></td>
<td><strong>49,194</strong></td>
</tr>
<tr>
<td>Private Use</td>
<td>5,632</td>
<td>6,400</td>
<td>6,400</td>
<td>6,400</td>
<td>6,400</td>
</tr>
<tr>
<td>Surrogates</td>
<td>2,048</td>
<td>2,048</td>
<td>2,048</td>
<td>2,048</td>
<td>2,048</td>
</tr>
<tr>
<td>Controls</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>Not Characters</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
D.2 Changes from Unicode Version 2.0 to Version 2.1

New Characters Added
Two new characters were added to the Unicode Standard, Version 2.1:

U+20AC euro sign
U+FFFC object replacement character

Character Semantics Changes
Significant clarifications or modifications to character semantics include the following:

Apostrophe. Because the character U+0027 apostrophe is very ambiguous, the preferred character for apostrophe was documented as either U+02BC modifier letter apostrophe or U+2019 right single quotation mark.

Bidirectional Properties. Certain characters were given new bidirectional properties definitions. U+0026 ampersand and U+0040 commercial at were changed from left to right (L) to other neutral (ON). U+002E full stop and U+2007 figure space were changed from European number separator (ES) to common number separator (CS).

Identifier Syntax. Corrections were made to the implementation guidelines for identifier syntax, adding or removing a few characters from the recommended set for inclusion in identifiers.

Mathematical and Letter Properties. A number of characters were given the mathematical property; see the list found in Unicode Technical Report #8, “The Unicode Standard, Version 2.1,” on the CD-ROM or the up-to-date version on the Unicode Web site for details. Two characters, U+02BC modifier letter apostrophe and U+055A armenian apostrophe, were removed from the list of the letter property.

Changes Affecting Conformance
Overall Unicode conformance criteria as described in Chapter 3 of Version 2.0 were unchanged. Aspects of the Unicode bidirectional algorithm were modified, Hangul syllable decompositions were clarified, and some normative character property values were changed. For details of these changes see Unicode Technical Report #8, “The Unicode Standard, Version 2.1,” on the CD-ROM or the up-to-date version on the Unicode Web site.
New Characters Added

In Version 3.0 of the Unicode Standard, 10,307 new characters have been added, as shown in Table D-3.

### Table D-3. New Characters Added

<table>
<thead>
<tr>
<th>Allocation</th>
<th>Count</th>
<th>Character Name</th>
</tr>
</thead>
</table>
| 01F6..01F9 | 4     | LATIN CAPITAL LETTER HWAIR  
|           |       | LATIN CAPITAL LETTER WYNN  
|           |       | LATIN CAPITAL LETTER N WITH GRAVE  
|           |       | LATIN SMALL LETTER N WITH GRAVE |
| 0218..021F | 8     | LATIN CAPITAL LETTER S WITH COMMA BELOW  
|           |       | LATIN SMALL LETTER S WITH COMMA BELOW  
|           |       | LATIN CAPITAL LETTER T WITH COMMA BELOW  
|           |       | LATIN SMALL LETTER T WITH COMMA BELOW  
|           |       | LATIN CAPITAL LETTER YOGH  
|           |       | LATIN SMALL LETTER YOGH  
|           |       | LATIN CAPITAL LETTER H WITH CARON  
|           |       | LATIN SMALL LETTER H WITH CARON |
| 0222..0233 | 18    | LATIN CAPITAL LETTER OU  
|           |       | LATIN SMALL LETTER OU  
|           |       | LATIN CAPITAL LETTER Z WITH HOOK  
|           |       | LATIN SMALL LETTER Z WITH HOOK  
|           |       | LATIN CAPITAL LETTER A WITH DOT ABOVE  
|           |       | LATIN SMALL LETTER A WITH DOT ABOVE  
|           |       | LATIN CAPITAL LETTER E WITH CEDILLA  
|           |       | LATIN SMALL LETTER E WITH CEDILLA  
|           |       | LATIN CAPITAL LETTER O WITH DIAERESIS AND MACRON  
|           |       | LATIN SMALL LETTER O WITH DIAERESIS AND MACRON  
|           |       | LATIN CAPITAL LETTER O WITH TILDE AND MACRON  
|           |       | LATIN SMALL LETTER O WITH TILDE AND MACRON  
|           |       | LATIN CAPITAL LETTER O WITH DOT ABOVE  
|           |       | LATIN SMALL LETTER O WITH DOT ABOVE  
|           |       | LATIN CAPITAL LETTER O WITH DOT ABOVE AND MACRON  
|           |       | LATIN SMALL LETTER O WITH DOT ABOVE AND MACRON  
|           |       | LATIN CAPITAL LETTER Y WITH MACRON  
|           |       | LATIN SMALL LETTER Y WITH MACRON |
| 02A9..02AD | 5     | LATIN SMALL LETTER FENG DIGRAPH  
|           |       | LATIN SMALL LETTER LS DIGRAPH  
|           |       | LATIN SMALL LETTER LZ DIGRAPH  
|           |       | LATIN LETTER BILABIAL PERCUSSIVE  
|           |       | LATIN LETTER BIDENTAL PERCUSSIVE |
| 02DF, 02EA..02EE | 6 | MODIFIER LETTER CROSS ACCENT  
|              |       | MODIFIER LETTER YIN DEPARTING TONE MARK  
|              |       | MODIFIER LETTER YANG DEPARTING TONE MARK  
|              |       | MODIFIER LETTER VOICING  
|              |       | MODIFIER LETTER UNASPIRATED  
|              |       | MODIFIER LETTER DOUBLE APOSTROPHE |
### Table D-3. New Characters Added (Continued)

<table>
<thead>
<tr>
<th>Allocation</th>
<th>Count</th>
<th>Character Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>0346..034E, 0362</td>
<td>10</td>
<td>COMBINING BRIDGE ABOVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COMBINING EQUALS SIGN BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COMBINING DOUBLE VERTICAL LINE BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COMBINING LEFT ANGLE BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COMBINING NOT TILDE ABOVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COMBINING HOMOTHETIC ABOVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COMBINING ALMOST EQUAL TO ABOVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COMBINING LEFT RIGHT ARROW BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COMBINING UPWARDS ARROW BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COMBINING DOUBLE RIGHTWARDS ARROW BELOW</td>
</tr>
<tr>
<td>03D7</td>
<td>1</td>
<td>GREEK KAI SYMBOL</td>
</tr>
<tr>
<td>03DB, 03DD, 03DF, 03E1</td>
<td>4</td>
<td>GREEK SMALL LETTER STIGMA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GREEK SMALL LETTER DIGAMMA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GREEK SMALL LETTER KOPPA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GREEK SMALL LETTER SAMPI</td>
</tr>
<tr>
<td>0400, 040D, 0450, 045D</td>
<td>4</td>
<td>CYRILLIC CAPITAL LETTER IE WITH GRAVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CYRILLIC SMALL LETTER IE WITH GRAVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CYRILLIC CAPITAL LETTER II WITH GRAVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CYRILLIC SMALL LETTER II WITH GRAVE</td>
</tr>
<tr>
<td>0488..0489</td>
<td>2</td>
<td>CYRILLIC HUNDRED THOUSANDS SIGN</td>
</tr>
<tr>
<td>048E..048F, 04EC..04ED</td>
<td>4</td>
<td>CYRILLIC CAPITAL LETTER ER WITH TICK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CYRILLIC SMALL LETTER ER WITH TICK</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CYRILLIC CAPITAL LETTER E WITH DIAERESIS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CYRILLIC SMALL LETTER E WITH DIAERESIS</td>
</tr>
<tr>
<td>058A</td>
<td>1</td>
<td>ARMENIAN HYPHEN</td>
</tr>
<tr>
<td>0653..0655, 0688..0689, 06BF, 06CF, 06FA..06FE</td>
<td>12</td>
<td>ARABIC MADDAH ABOVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC HAMZA ABOVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC HAMZA BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC LETTER LAM WITH THREE DOTS BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC LETTER NOON WITH DOT BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC LETTER TCHEH WITH DOT ABOVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC LETTER WAW WITH DOT ABOVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC LETTER SHEEN WITH DOT BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC LETTER DAD WITH DOT BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC LETTER GHAIN WITH DOT BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC SIGN SINDHI AMPERSAND</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ARABIC SIGN SINDHI POSTPOSITION MEN</td>
</tr>
<tr>
<td>0700..074A</td>
<td>71</td>
<td>Syriac</td>
</tr>
<tr>
<td>0780..07B1</td>
<td>50</td>
<td>Thaana</td>
</tr>
<tr>
<td>0D80..0DFF</td>
<td>80</td>
<td>Sinhala</td>
</tr>
<tr>
<td>0F6A, 0F96, 0FAE..0FCF</td>
<td>25</td>
<td>Tibetan Extensions</td>
</tr>
<tr>
<td>1000..1059</td>
<td>78</td>
<td>Myanmar</td>
</tr>
<tr>
<td>1200..137F</td>
<td>346</td>
<td>Ethiopic</td>
</tr>
<tr>
<td>13A0..13FF</td>
<td>85</td>
<td>Cherokee</td>
</tr>
<tr>
<td>1401..167F</td>
<td>630</td>
<td>Canadian Syllabics</td>
</tr>
<tr>
<td>1680..169F</td>
<td>29</td>
<td>Ogham</td>
</tr>
<tr>
<td>16A0..16FF</td>
<td>81</td>
<td>Runic</td>
</tr>
<tr>
<td>1780..17E9</td>
<td>103</td>
<td>Khmer</td>
</tr>
<tr>
<td>1800..18A9</td>
<td>155</td>
<td>Mongolian</td>
</tr>
<tr>
<td>Allocation</td>
<td>Count</td>
<td>Character Name</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| 202F, 2048..204D | 7     | NARROW NO-BREAK SPACE            
QUESTION EXCLAMATION MARK              
EXCLAMATION QUESTION MARK              
TIRONIAN SIGH ET                     
REVERSED PILCROW SIGN                
BLACK LEFTWARDS BULLET               
BLACK RIGHTWARDS BULLET               |
| 20AD..20AF   | 3     | KIP SIGN                         
TUGRIK SIGN                          
DRACHMA SIGN                         |
| 20E2..20E3   | 2     | COMBINING ENCLOSING SCREEN       
COMBINING ENCLOSING KEYCAP           |
| 2139..213A   | 2     | INFORMATION SOURCE               
ROTATED CAPITAL Q                    |
| 2183         | 1     | ROMAN NUMERAL REVERSED ONE HUNDRED |
| 21EB..21F3   | 9     | UPWARDS WHITE ARROW ON PEDESTAL  
UPWARDS WHITE ARROW ON PEDESTAL WITH HORIZONTAL BAR  
UPWARDS WHITE ARROW ON PEDESTAL WITH VERTICAL BAR  
UPWARDS WHITE DOUBLE ARROW           
UPWARDS WHITE DOUBLE ARROW ON PEDESTAL  
RIGHTWARDS WHITE ARROW FROM WALL     
NORTH WEST ARROW TO CORNER           
SOUTH EAST ARROW TO CORNER           
UP DOWN WHITE ARROW                   |
| 2301, 237B, 237D, 237E..237F | 5     | ELECTRIC ARROW                   
NOT CHECK MARK                       
SHOULDERED OPEN BOX                  
BELL SYMBOL                          
VERTICAL LINE WITH MIDDLE DOT        |
| 2380..238C   | 13    | INSERTION SYMBOL                 
CONTINUOUS UNDERLINE SYMBOL          
DISCONTINUOUS UNDERLINE SYMBOL       
EMPHASIS SYMBOL                      
COMPOSITION SYMBOL                   
WHITE SQUARE WITH CENTRE VERTICAL LINE  
ENTER SYMBOL                         
ALTERNATIVE KEY SYMBOL               
HELM SYMBOL                          
CIRCLED HORIZONTAL BAR WITH NOTCH    
CIRCLED TRIANGLE DOWN                 
BROKEN CIRCLE WITH NORTHWEST ARROW   
UNDO SYMBOL                          |
| 238D..239A   | 14    | MONOSTABLE SYMBOL                
HYSTERESIS SYMBOL                    
OPEN-CIRCUIT-OUTPUT H-TYPE SYMBOL    
OPEN-CIRCUIT-OUTPUT L-TYPE SYMBOL    
PASSIVE-PULL-DOWN-OUTPUT SYMBOL     
PASSIVE-PULL-UP-OUTPUT SYMBOL       
DIRECT CURRENT SYMBOL FORM TWO       
SOFTWARE-FUNCTION SYMBOL             
APL FUNCTIONAL SYMBOL QUAD          
DECIMAL SEPARATOR KEY SYMBOL         
PREVIOUS PAGE                         
NEXT PAGE                             
PRINT SCREEN SYMBOL                  
CLEAR SCREEN SYMBOL                   |
| 2425..2426   | 2     | SYMBOL FOR DELETE FORM TWO       
SYMBOL FOR SUBSTITUTE FORM TWO       |
### Table D-3. New Characters Added (Continued)

<table>
<thead>
<tr>
<th>Allocation</th>
<th>Count</th>
<th>Character Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>25F0..25F7</td>
<td>8</td>
<td>WHITE SQUARE WITH UPPER LEFT QUADRANT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WHITE SQUARE WITH LOWER LEFT QUADRANT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WHITE SQUARE WITH LOWER RIGHT QUADRANT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WHITE SQUARE WITH UPPER RIGHT QUADRANT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WHITE CIRCLE WITH UPPER LEFT QUADRANT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WHITE CIRCLE WITH LOWER LEFT QUADRANT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WHITE CIRCLE WITH LOWER RIGHT QUADRANT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WHITE CIRCLE WITH UPPER RIGHT QUADRANT</td>
</tr>
<tr>
<td>2619, 2670..2671</td>
<td>3</td>
<td>REVERSED ROTATED FLORAL HEART BULLET</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WEST SYRIAC CROSS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EAST SYRIAC CROSS</td>
</tr>
<tr>
<td>2800..28FF</td>
<td>256</td>
<td>Braille Pattern Symbols</td>
</tr>
<tr>
<td>2E80..2EF3</td>
<td>115</td>
<td>CJK Radicals Supplement</td>
</tr>
<tr>
<td>2F00..2FD5</td>
<td>214</td>
<td>KangXi radicals</td>
</tr>
<tr>
<td>2FF0..2FFB</td>
<td>12</td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER LEFT TO RIGHT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER ABOVE TO BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER LEFT TO MIDDLE AND RIGHT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER ABOVE TO MIDDLE AND BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER FULL SURROUND</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM ABOVE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM BELOW</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM LEFT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM UPPER LEFT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM UPPER RIGHT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER SURROUND FROM LOWER LEFT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC DESCRIPTION CHARACTER OVERLAID</td>
</tr>
<tr>
<td>3038..303A, 303E</td>
<td>4</td>
<td>HANGZHOU NUMERAL TEN</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HANGZHOU NUMERAL TWENTY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HANGZHOU NUMERAL THIRTY</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IDEOGRAPHIC VARIATION INDICATOR</td>
</tr>
<tr>
<td>31A0..31B7</td>
<td>24</td>
<td>Extended Bopomofo</td>
</tr>
<tr>
<td>3400..4DFF</td>
<td>6,582</td>
<td>CJK Unified Ideographs, Extension A</td>
</tr>
<tr>
<td>A000..A48C</td>
<td>1,165</td>
<td>YI</td>
</tr>
<tr>
<td>A490..A4C1</td>
<td>50</td>
<td>YI radicals</td>
</tr>
<tr>
<td>FB1D</td>
<td>1</td>
<td>HEBREW LETTER YOD WITH HIRIQ</td>
</tr>
<tr>
<td>FFF9..FFF8</td>
<td>3</td>
<td>INTERLINEAR ANNOTATION ANCHOR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>INTERLINEAR ANNOTATION SEPARATOR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>INTERLINEAR ANNOTATION TERMINATOR</td>
</tr>
</tbody>
</table>

### Character Semantics Changes

Significant clarifications or modifications to character semantics include the following:

- **Bidirectional Properties**: Bidirectional properties were made more consistent with the general category property, and new bidirectional properties were created. See Section 3.12, Bidirectional Behavior.

- **Byte Order Mark**: The use of the byte order mark with transformation formats was clarified. See Section 3.8, Transformations.
Changes from Unicode Version 2.0 D.3 Changes from Unicode Version 2.1 to Version 3.0

Capital Letters with Iota Adscript. The representative glyphs, semantics, case mappings, and decompositions have been revised to make their handling more consistent.

Case. Case properties have been extended for those situations where there is a mapping to multiple characters and where case is locale-dependent. See the SpecialCasing.txt file on the CD-ROM.

Combining Classes. These classes were updated significantly to resolve problems of normalization and decomposition for Indic scripts in particular. See Table 4-3, Combining Classes.

Decompositions. Unicode character decompositions have been significantly updated to fix errors in the original assignments, to allow correct collation weighting, and to make decompositions consistent for normalization.

Eyelash Ra. Consonant RA rules have been updated and expanded. See rules R5 and R5a in Section 9.1, Devanagari.

Figure Space. U+2007 figure space is no longer treated like a numeric separator for purposes of bidirectional layout. See Section 6.1, General Punctuation, for a description of its tabular width characteristics.

General Category. A series of General Category changes were made to assist the convergence of the Unicode definition of identifier with ISO TR 10176.

Identifier Syntax. The list of recommended characters for computer language identifiers was corrected again, and the syntax for identifiers was further simplified. (See Section 5.16, Identifiers.)

Layout Controls. The description of layout controls was enhanced to include the behavior of U+00A0 no-break space, U+00AD soft hyphen, and zero width spaces. See Section 13.2, Layout Controls.

Line and Paragraph Separators. Use of line and paragraph separators is clarified in Section 3.12, Bidirectional Behavior, and in Section 13.1, Control Codes.

Newlines. Line-handling characteristics have been documented more fully for Unicode environments. Discussions on CR and LF can be found in Section 3.12, Bidirectional Behavior; Section 5.9, Line Handling; Section 13.1, Control Codes; and in Unicode Technical Report #13, “Unicode Newline Guidelines,” on the CD-ROM or the up-to-date version on the Unicode Web site.

Quotation Marks. Two new punctuation categories, Pi and Pf, were created for initial and final quotes. The use of these categories and language-based usage of quotation marks is clarified in Section 6.1, General Punctuation.

Script Capital p. The Weierstrass elliptic function symbol, U+2118 script capital p, actually has the form of a lowercase calligraphic p despite the use of capital in its name.

Symmetric Swapping. The symmetric swapping value for guillemets was corrected.

Tibetan. Semantics of many Tibetan characters have been clarified or revised. See Section 9.13, Tibetan.

Tilde. The use of U+007E tilde as a spacing clone of combining tilde and as a regular character is described in Section 6.1, General Punctuation.

Changes Affecting Conformance

Conformance clauses, definitions, and explanatory text were added for handling Unicode Transformation Formats. The Unicode bidirectional algorithm rules were clarified and expanded, and new bidirectional character properties were documented. Other normative
D.3 Changes from Unicode Version 2.1 to Version 3.0

Changes from Unicode Version 2.0

character property values were changed; see the Unicode Character Database found on the CD-ROM for details.

**Unicode Technical Reports**

- UTR #11: East Asian Width, Version 5.0
- UTR #13: Unicode Newline Guidelines, Version 5.0
- UTR #14: Line Breaking Properties, Version 6.0
- UTR #15: Unicode Normalization Forms, Version 18.0