2.0 Changes in Unicode 1.0

As discussed in Volumes 1 and 2, small changes have been made to Unicode 1.0 in order to align it with the international character encoding standard, ISO/IEC 10646-1. In order to expedite use of Unicode in the interim, the Unicode Consortium issued an intermediate version, Unicode 1.0.1, which consisted of Unicode 1.0 modified by the changes necessary to make the character codes a proper subset of ISO/IEC 10646-1. The text of the 1.0.1 document is subsumed by this document.

2.1 Final Zone Allocations

ISO/IEC 10646-1 is organized into zones, which were reallocated in the published version. These zone reallocations do not affect any allocated Unicode 1.0 characters, but do affect the Private Use Area.

2.1.1 Unicode Allocation

<table>
<thead>
<tr>
<th>Range</th>
<th>Cells</th>
<th>Name/Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+0000 → U+4DFF</td>
<td>19,968</td>
<td>A-ZONE Alphabets, syllabaries, symbols (the 65 control codes are excluded)</td>
</tr>
<tr>
<td>U+4E00 → U+9FFF</td>
<td>20,992</td>
<td>I-ZONE Ideographs</td>
</tr>
<tr>
<td>U+A000 → U+DFFF</td>
<td>16,384</td>
<td>O-ZONE Reserved for future assignment</td>
</tr>
<tr>
<td>U+E000 → U+FFF</td>
<td>8,192</td>
<td>R-ZONE Restricted use (FFFFE and FFFF are excluded)</td>
</tr>
</tbody>
</table>

2.1.2 R-ZONE Allocation

<table>
<thead>
<tr>
<th>Range</th>
<th>Cells</th>
<th>Name/Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+E070 → U+F8FF</td>
<td>6,400</td>
<td>Private Use Area (Corporate Use Zone starts at F8FF allocating downwards; End User Zone starts at E000 allocating upwards)</td>
</tr>
<tr>
<td>U+F900 → U+FFE</td>
<td>1,776</td>
<td>Compatibility Zone (including presentation forms)</td>
</tr>
<tr>
<td>U+FFF0 → U+FFF</td>
<td>16</td>
<td>Specials (FFFFE and FFFF are not character codes, and are excluded)</td>
</tr>
</tbody>
</table>

2.2 Characters deleted and withdrawn for further study:

2.2.1 Groups of characters deleted

<table>
<thead>
<tr>
<th>Range</th>
<th>Group Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+0E70 → U+0E74</td>
<td>Thai Phonetic Order Vowel signs</td>
</tr>
<tr>
<td>U+0EFO → U+0E4</td>
<td>Lao Phonetic Order Vowel signs</td>
</tr>
<tr>
<td>U+1000 → U+104C</td>
<td>Tibetan script</td>
</tr>
</tbody>
</table>

2.2.2 Individual characters deleted

<table>
<thead>
<tr>
<th>Code</th>
<th>Image</th>
<th>Old Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+03DB</td>
<td>Ξ</td>
<td>GREEK SMALL LETTER STIGMA</td>
</tr>
<tr>
<td>U+03DD</td>
<td>Γ</td>
<td>GREEK SMALL LETTER DIGAMMA</td>
</tr>
<tr>
<td>U+03DF</td>
<td>Ω</td>
<td>GREEK SMALL LETTER KOPPA</td>
</tr>
<tr>
<td>U+03E1</td>
<td>Β</td>
<td>GREEK SMALL LETTER SAMPI</td>
</tr>
<tr>
<td>U+2300</td>
<td>-</td>
<td>APL COMPOSE</td>
</tr>
<tr>
<td>U+2301</td>
<td>[</td>
<td>APL OUT</td>
</tr>
</tbody>
</table>

ΈΠ U+0A3D GURMUKHI SIGN AVAGRAHA was retained in Unicode 1.0.1 but is now deleted in 1.1
2.3 Characters unified

<table>
<thead>
<tr>
<th>From</th>
<th>With</th>
<th>Image</th>
<th>Old Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+0371</td>
<td>U+0314</td>
<td></td>
<td>GREEK NON-SPACING DASIA PNEUMATA</td>
</tr>
<tr>
<td>U+0372</td>
<td>U+0313</td>
<td></td>
<td>GREEK NON-SPACING PSILI PNEUMATA</td>
</tr>
<tr>
<td>U+0384</td>
<td>U+030D</td>
<td></td>
<td>GREEK NON-SPACING TONOS</td>
</tr>
<tr>
<td>U+04C5</td>
<td>U+049A</td>
<td>K</td>
<td>CYRILLIC CAPITAL LETTER KA OGOINEK</td>
</tr>
<tr>
<td>U+04C6</td>
<td>U+049B</td>
<td>K</td>
<td>CYRILLIC SMALL LETTER KA OGOINEK</td>
</tr>
<tr>
<td>U+04C9</td>
<td>U+04B2</td>
<td>X</td>
<td>CYRILLIC CAPITAL LETTER KHA OGOINEK</td>
</tr>
<tr>
<td>U+04CA</td>
<td>U+04B3</td>
<td>X</td>
<td>CYRILLIC SMALL LETTER KHA OGOINEK</td>
</tr>
<tr>
<td>U+3004</td>
<td>U+4edd</td>
<td></td>
<td>IDEOGRAPHIC DITTO MARK</td>
</tr>
</tbody>
</table>

2.4 Characters moved

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Image</th>
<th>Old Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+0370</td>
<td>U+0345</td>
<td></td>
<td>GREEK NON-SPACING IOTA BELOW</td>
</tr>
<tr>
<td>U+0385</td>
<td>U+0344</td>
<td></td>
<td>GREEK NON-SPACING DIAERESIS TONOS</td>
</tr>
<tr>
<td>U+03D7</td>
<td>U+037E</td>
<td>;</td>
<td>GREEK QUESTION MARK</td>
</tr>
<tr>
<td>U+03D8</td>
<td>U+0374</td>
<td>,</td>
<td>GREEK UPPER NUMERAL SIGN</td>
</tr>
<tr>
<td>U+03D9</td>
<td>U+0375</td>
<td>,</td>
<td>GREEK LOWER NUMERAL SIGN</td>
</tr>
<tr>
<td>U+03F3</td>
<td>U+0384</td>
<td></td>
<td>GREEK SPACING TONOS</td>
</tr>
<tr>
<td>U+03F4</td>
<td>U+0385</td>
<td></td>
<td>GREEK SPACING DIAERESIS TONOS</td>
</tr>
<tr>
<td>U+03F5</td>
<td>U+037A</td>
<td></td>
<td>GREEK SPACING IOTA BELOW</td>
</tr>
<tr>
<td>U+05F5</td>
<td>U+F1EB</td>
<td>☞</td>
<td>HEBREW POINT VARICA</td>
</tr>
<tr>
<td>U+32FF</td>
<td>U+3004</td>
<td>☞</td>
<td>JAPANESE INDUSTRIAL STANDARD SYMBOL</td>
</tr>
</tbody>
</table>

2.5 Character blocks rearranged

The explicit list is in Appendix I: Unicode 1.1 Character List, p. 43.

<table>
<thead>
<tr>
<th>Range</th>
<th>Group Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+3200 → U+32FF</td>
<td>Circed Katakana: The 1.1 characters are arranged in modern order: i.e., A, I, U, E, O, KA, KI,...</td>
</tr>
<tr>
<td>U+FE80 → U+FEFC</td>
<td>Basic glyphs for Arabic language: The 1.1 character shapes are arranged in different order: Isolate, Final, Initial, Medial</td>
</tr>
</tbody>
</table>

2.6 Character mapping changed

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Image</th>
<th>XJIS</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>U+00AD</td>
<td>U+2010</td>
<td>-</td>
<td>815D</td>
<td>JIS HYPHEN</td>
</tr>
<tr>
<td>U+20DD</td>
<td>U+25EF</td>
<td>☞</td>
<td>81FC</td>
<td>JIS COMPOSITION CIRCLE</td>
</tr>
</tbody>
</table>

2.7 Control Characters

ISO/IEC 10646-1 forbids the use of the C1 control characters from U+0080 → U+009F. Instead, it reserves those code locations, and requires use of an ESC FE sequence instead (see ISO/IEC 6429). This does not affect Unicode implementations because Unicode does not specify the use of control codes (beyond how to express them in a 16-bit form), and considers such specification to be subject to other protocols.