

Appendix D

Cumulative Changes

D.1 Versions of the Unicode Standard

The Unicode Technical Committee has accepted changes to the Unicode Standard in order to maintain a single numerical encoding provided by the Unicode Standard and the standard ISO/IEC 10646-1:1993. There have been three formally numbered versions of the Unicode Standard, relating to versions of ISO 10646 as shown in Table D-1.

Table D-1. Versions of the Unicode Standard

Year	Version	Published	Content Relation to ISO 10646
1991	Unicode 1.0	Addison-Wesley	Basis of Draft-2 ISO 10646
1993	Unicode 1.1	Unicode Tech Report #4	Matches ISO 10646
1996	Unicode 2.0	Addison-Wesley	Matches ISO 10646 plus amendments

The Unicode Standard since version 1.1 has been kept in synch with ISO 10646 so that the two standards match precisely in character content and naming; the history of this relationship is given in *Appendix C, Relationship to ISO/IEC 10646*. This appendix enumerates the specific character content, name, and semantic changes made from Unicode Standard, Version 1.0 to Version 1.1 and from Version 1.1 to Version 2.0 to maintain alignment with ISO 10646.

D.2 Changes from Unicode 1.0 to 1.1

Areas Redefined

Private Use Area. In the Unicode Standard, Version 1.0, the Private Use Area ranged from U+E800 → U+FDFF (5,632 code positions). In the Unicode Standard, Version 1.1, this area was enlarged and shifted to the range U+E000 → U+F8FF (6,400 code positions).

Characters Removed

Four Greek letters were removed as part of a realignment of the Greek block; these case forms were judged inappropriate:

U+03DB GREEK SMALL LETTER STIGMA
U+03DD GREEK SMALL LETTER DIGAMMA
U+03DF GREEK SMALL LETTER KOPPA
U+03E1 GREEK SMALL LETTER SAMPI

One Gurmukhi sign was removed:

U+0A3D GURMUKHI SIGN AVAGRAHA

Five Thai phonetic order vowel signs were removed at the request of the Thai National Body as part of the ISO/IEC 10646 merger process in order to conduct further study on their utilization. Consequently, the alternate ordering described in the Unicode Standard, Version 1.0 no longer applies:

```
U+0E70 THAI PHONETIC ORDER VOWEL SIGN SARA E
U+0E71 THAI PHONETIC ORDER VOWEL SIGN SARA AE
U+0E72 THAI PHONETIC ORDER VOWEL SIGN SARA O
U+0E73 THAI PHONETIC ORDER VOWEL SIGN SARA MAI
      MUAN
U+0E74 THAI PHONETIC ORDER VOWEL SIGN SARA MAI
      MALAI
```

Five Lao phonetic order vowel signs were removed analogously with those in Thai:

```
U+0EF0 LAO PHONETIC ORDER VOWEL SIGN E
U+0EF1 LAO PHONETIC ORDER VOWEL SIGN EI
U+0EF2 LAO PHONETIC ORDER VOWEL SIGN O
U+0EF3 LAO PHONETIC ORDER VOWEL SIGN AY
U+0EF4 LAO PHONETIC ORDER VOWEL SIGN AI
```

The entire Tibetan block of the Unicode Standard, Version 1.0, comprising 71 assigned characters (U+1000 → U+105F), was removed for further study. (Tibetan has returned in the Unicode Standard, Version 2.0, occupying a reorganized block in the range U+0F00 → U+0FBF, see additions below.):

Two APL symbols were removed as part of a realignment of the APL group:

```
U+2300 APL COMPOSE OPERATOR1
U+2301 APL OUT
```

And note removal of a glyph shown by error in the Unicode Standard, Version 1.0 chart:

```
U+27B0 (never was a character)
```

Characters Unified

Three Greek letters were unified (that is, moved to merge with another existing character) as part of a realignment of the Greek block. Four Cyrillic letters were unified as being variants. One CJK punctuation symbol was unified as being an ideographic character. See Table D-2.

Table D-2. Characters Unified

From	To/With	1.0 Name
U+0371	U+0314	GREEK NON-SPACING DASIA PNEUMATA
U+0372	U+0313	GREEK NON-SPACING PSILI PNEUMATA
U+0384	U+030D	GREEK NON-SPACING TONOS ^a
U+04C5	U+049A	CYRILLIC CAPITAL LETTER KA OGONEK
U+04C6	U+049B	CYRILLIC SMALL LETTER KA OGONEK
U+04C9	U+04B2	CYRILLIC CAPITAL LETTER KHA OGONEK
U+04CA	U+04B3	CYRILLIC SMALL LETTER KHA OGONEK
U+3004	U+4EDD	IDEOGRAPHIC DITTO MARK ^b

- This code has been assigned to a moved character, GREEK TONOS (renamed from GREEK SPACING TONOS).
- This code has been assigned to a moved character, JAPANESE INDUSTRIAL STANDARD SYMBOL.

1. This code has been assigned to a new character, DIAMETER SIGN.

Characters Moved

Eight Greek characters were moved as part of a realignment of the Greek block. One Hebrew mark was moved to the Alphabetic Presentation Forms block. One Japanese symbol was moved. See Table D-3.

Table D-3. Characters Moved

From	To	1.0 Name
U+0370	U+0345	GREEK NON-SPACING IOTA BELOW
U+0385	U+0344	GREEK NON-SPACING DIAERESIS TONOS
U+03D7	U+037E	GREEK QUESTION MARK
U+03D8	U+0374	GREEK UPPER NUMERAL SIGN
U+03D9	U+0375	GREEK LOWER NUMERAL SIGN
U+03F3	U+0384	GREEK SPACING TONOS ^a
U+03F4	U+0385	GREEK SPACING DIAERESIS TONOS ^b
U+03F5	U+037A	GREEK SPACING IOTA BELOW
U+05F5	U+FB1E	HEBREW POINT VARİKA
U+32FF	U+3004	JAPANESE INDUSTRIAL STANDARD SYMBOL ^c

- This replaces U+0384 GREEK NON-SPACING TONOS; the code U+03F3 has been assigned to a new character, GREEK LETTER YOT.
- This replaces U+0385 GREEK NON-SPACING DIAERESIS TONOS; the code U+03F4 has been assigned to a moved character, GREEK DIALYTIKA TONOS (renamed from GREEK SPACING DIAERESIS TONOS).
- This replaces U+3004 IDEOGRAPHIC DITTO MARK.

All 47 circled Katakana characters (U+32D0 → U+32FE) were reordered from I-RO-HA order to A-I-U-E-O order.

All 69 Arabic contextual-variant glyphs in the Arabic Presentation Forms-B block (U+FE8A → U+FEF4) were reordered from INITIAL-MEDIAL-FINAL order to FINAL-INITIAL-MEDIAL order.

Code Values Whose Assignment Has Changed

The information in Tables D-4 and D-5 is not additional to the previous categories; it can be derived from the foregoing. However, because of its importance it is presented here in summary form. These are all the code values for which the Unicode Standard, Version 1.0 and the Unicode Standard, Version 1.1 have *different characters* at the same codepoint.

Table D-4. Changed Assignments

Code	Unicode 1.0 Character	Unicode 1.1 Character
U+0384	GREEK NON-SPACING TONOS	GREEK TONOS
U+0385	GREEK NON-SPACING DIAERESIS TONOS	GREEK DIALYTIKA TONOS
U+03F3	GREEK SPACING TONOS	GREEK LETTER YOT
U+2300	APL COMPOSE OPERATOR	DIAMETER SIGN
U+3004	IDEOGRAPHIC DITTO MARK	JAPANESE INDUSTRIAL STANDARD SYMBOL

Table D-5. Reordered Character Groups

Code Range	Character Group
U+32D0 → U+32FE	Circled Katakana
U+FE8A → U+FEF4	Arabic contextual-variant glyphs

Apart from the 5 individual codepoints just enumerated plus two groups totaling 116 others, the following conclusion can be drawn: If the Unicode Standard, Version 1.0 and the

Unicode Standard, Version 1.1 both contain a given codepoint but the names attached to that codepoint differ, then this difference represents merely a *name change* for the same character assignment, not a differing character assignment.

New Characters Added

In the Unicode Standard, Version 1.1, a total of 4,306 Korean Hangul “Supplementary Syllables” plus 1,964 other items from drafts of ISO 10646 were annexed into code positions that had been unassigned in the Unicode Standard, Version 1.0. The Hangul supplementary syllables have since been removed in the Unicode Standard, Version 2.0; most of the remaining additions from the Unicode Standard, Version 1.0 to the Unicode Standard, Version 1.1 are presentation forms whose use as characters is discouraged in the Unicode Standard. These additions are listed below by the Unicode Standard, Version 2.0 block; in this listing, the relatively few additions that constitute substantive characters are enumerated explicitly.

In block Latin Extended-A, 1 additional precomposed diacritic combination.

In block Latin Extended-B, 35 additional precomposed diacritic combinations.

In block Combining Diacritical Marks, 6 additions:

U+0342 COMBINING GREEK PERISPOMENI
 U+0343 COMBINING GREEK KORONIS
 U+0344 COMBINING GREEK DIALYTIKA TONOS
 U+0345 COMBINING GREEK YPOGEGRAMMENI
 U+0360 COMBINING DOUBLE TILDE
 U+0361 COMBINING DOUBLE INVERTED BREVE

In block Greek, 5 additions:

U+0374 GREEK NUMERAL SIGN
 U+0375 GREEK LOWER NUMERAL SIGN
 U+037A GREEK YPOGEGRAMMENI
 U+037E GREEK QUESTION MARK
 U+0387 GREEK ANO TELEIA

In block Cyrillic, 38 additional precomposed diacritic combinations.

In block Armenian, 1 additional ligature.

In block Arabic, 25 additions, largely Koranic reading symbols:

U+066D ARABIC FIVE POINTED STAR
 U+06D6 ARABIC SMALL HIGH LIGATURE SAD WITH LAM
 WITH ALEF MAKSURA
 U+06D7 ARABIC SMALL HIGH LIGATURE QAF WITH LAM
 WITH ALEF MAKSURA
 U+06D8 ARABIC SMALL HIGH MEEM INITIAL FORM
 U+06D9 ARABIC SMALL HIGH LAM ALEF
 U+06DA ARABIC SMALL HIGH JEEM
 U+06DB ARABIC SMALL HIGH THREE DOTS
 U+06DC ARABIC SMALL HIGH SEEN
 U+06DD ARABIC END OF AYAH
 U+06DE ARABIC START OF RUB EL HIZB
 U+06DF ARABIC SMALL HIGH ROUNDED ZERO
 U+06E0 ARABIC SMALL HIGH UPRIGHT RECTANGULAR
 ZERO
 U+06E1 ARABIC SMALL HIGH DOTLESS HEAD OF KHAH
 U+06E2 ARABIC SMALL HIGH MEEM ISOLATED FORM
 U+06E3 ARABIC SMALL LOW SEEN
 U+06E4 ARABIC SMALL HIGH MADDA
 U+06E5 ARABIC SMALL WAW

U+06E6 ARABIC SMALL YEH
 U+06E7 ARABIC SMALL HIGH YEH
 U+06E8 ARABIC SMALL HIGH NOON
 U+06E9 ARABIC PLACE OF SAJDAH
 U+06EA ARABIC EMPTY CENTRE LOW STOP
 U+06EB ARABIC EMPTY CENTRE HIGH STOP
 U+06EC ARABIC ROUNDED HIGH STOP WITH FILLED
 CENTRE
 U+06ED ARABIC SMALL LOW MEEM

In block Gurmukhi, 1 addition:

U+0A4D GURMUKHI SIGN VIRAMA

In block Gujarati, 3 additions:

U+0A8D GUJARATI VOWEL CANDRA E
 U+0A91 GUJARATI VOWEL CANDRA O
 U+0AC9 GUJARATI VOWEL SIGN CANDRA O

In block Oriya, 1 addition:

U+0B56 ORIYA AI LENGTH MARK

The whole block Hangul Jamo, comprising 240 additions.

The whole block Latin Extended Additional, comprising 245 additional precomposed diacritic combinations.

The whole block Greek Extended, comprising 233 additional precomposed diacritic combinations.

In block General Punctuation, 3 additions:

U+203F UNDERTIE
 U+2045 LEFT SQUARE BRACKET WITH QUILL
 U+2046 RIGHT SQUARE BRACKET WITH QUILL

Also in block General Punctuation, 6 additional form-shaping controls.

In block Miscellaneous Technical, 9 additions:

U+232D CYLINDRICITY
 U+232E ALL AROUND-PROFILE
 U+232F SYMMETRY
 U+2330 TOTAL RUNOUT
 U+2331 DIMENSION ORIGIN
 U+2332 CONICAL TAPER
 U+2333 SLOPE
 U+2334 COUNTERBORE
 U+2335 COUNTERSINK

Also in block Miscellaneous Technical, 69 additional precomposed APL functional symbols.

In block Geometric Shapes, 1 addition:

U+25EF LARGE CIRCLE

In block CJK Symbols and Punctuation, 1 addition:

U+3037 IDEOGRAPHIC TELEGRAPH LINE FEED SEPARATOR SYMBOL

In block Katakana, 4 additions:

U+30F7 KATAKANA LETTER VA
 U+30F8 KATAKANA LETTER VI
 U+30F9 KATAKANA LETTER VE
 U+30FA KATAKANA LETTER VO

In block Enclosed CJK Letters and Months, 12 additional telegraph symbols for the months.

In block CJK Compatibility, 62 additions, mostly telegraph symbols.

The whole block Hangul Supplementary Syllables A, comprising 1,930 modern Korean Hangul syllables from the standard KS C 5659-1990—*removed in the Unicode Standard, Version 2.0.*

The whole block Hangul Supplementary Syllables B, comprising 2,376 old Korean Hangul syllables from the standard KS C 5657-1991—*removed in the Unicode Standard, Version 2.0.*

The whole block CJK Compatibility Ideographs, comprising 302 additional compatibility variants

The whole block Alphabetic Presentation Forms, comprising 57 additional presentation forms.

The whole block Arabic Presentation Forms-A, comprising 593 additional presentation forms.

The whole block Combining Half Marks, comprising 4 additional presentation forms

In block Halfwidth and Fullwidth Forms, 7 additional halfwidth forms.

Character Name Changes

Characters in the Unicode Standard and ISO 10646, except for CJK Ideographs and Hangul Syllables, are each given a unique name written in all uppercase, with at most a hyphen for punctuation. ISO 10646 names may also have appended a parenthesized comment, which are only informational. In the Unicode Standard, Version 1.1, the character names were adjusted to match the ISO 10646 names precisely. This involved changing the names of 1,890 characters, though in most cases the changes were only minor editorial variations.

The full list of editorial changes will not be enumerated here, as the full set of the Unicode Standard, Version 1.0 character names may be found on the Unicode CD-ROM as part of the *Unicode Character Database*. As was stated in the subsection “Code Values Whose Assignment Has Changed”, only 5 individual codepoints, plus two groups totaling 116 others, actually changed in character assignment from the Unicode Standard, Version 1.0 to the Unicode Standard, Version 1.1. In all other cases where the names attached to a codepoint differ, this difference represents an editorial change in the character name.

A few comments may be added concerning the nature of the editorial name changes. Some changes apply mainly or solely to a single block while others are applied throughout the names list. For example, the Unicode Standard, Version 1.0 has “PERIOD” where the Unicode Standard, Version 1.1 has “FULL STOP”, this change is made consistently in the whole names list, for example:

Code	1.0 Name	1.1 Name
002E	PERIOD	FULL STOP
0589	ARMENIAN PERIOD	ARMENIAN FULL STOP
06D4	ARABIC PERIOD	ARABIC FULL STOP
2488	DIGIT ONE PERIOD	DIGIT ONE FULL STOP

There are other cases of terminology change which apply to many characters but which do not map 100% systematically between the two versions. There are also some changes which are individual to particular single characters.

For reference, *Table D-6 1.1 Name Changes*, on page D-7 shows *all* the name differences, regardless of consistency, occurring within the first 256 codepoints:

Table D-6. 1.1 Name Changes

Code	Unicode 1.0	Unicode 1.1
0027	APOSTROPHE-QUOTE	APOSTROPHE
0028	OPENING PARENTHESIS	LEFT PARENTHESIS
0029	CLOSING PARENTHESIS	RIGHT PARENTHESIS
002E	PERIOD	FULL STOP
002F	SLASH	SOLIDUS
005B	OPENING SQUARE BRACKET	LEFT SQUARE BRACKET
005C	BACKSLASH	REVERSE SOLIDUS
005D	CLOSING SQUARE BRACKET	RIGHT SQUARE BRACKET
005E	SPACING CIRCUMFLEX	CIRCUMFLEX ACCENT
005F	SPACING UNDERSCORE	LOW LINE
0060	SPACING GRAVE	GRAVE ACCENT
007B	OPENING CURLY BRACKET	LEFT CURLY BRACKET
007C	VERTICAL BAR	VERTICAL LINE
007D	CLOSING CURLY BRACKET	RIGHT CURLY BRACKET
00A0	NON-BREAKING SPACE	NO-BREAK SPACE
00A6	BROKEN VERTICAL BAR	BROKEN BAR
00A8	SPACING DIAERESIS	DIAERESIS
00AB	LEFT POINTING GUILLEMET	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK
00AE	REGISTERED TRADE MARK SIGN	REGISTERED SIGN
00AF	SPACING MACRON	MACRON
00B1	PLUS-OR-MINUS SIGN	PLUS-MINUS SIGN
00B2	SUPERSCRIPIT DIGIT TWO	SUPERSCRIPIT TWO
00B3	SUPERSCRIPIT DIGIT THREE	SUPERSCRIPIT THREE
00B4	SPACING ACUTE	ACUTE ACCENT
00B6	PARAGRAPH SIGN	PILCROW SIGN
00B8	SPACING CEDILLA	CEDILLA
00B9	SUPERSCRIPIT DIGIT ONE	SUPERSCRIPIT ONE
00BB	RIGHT POINTING GUILLEMET	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK
00BC	FRACTION ONE QUARTER	VULGAR FRACTION ONE QUARTER
00BD	FRACTION ONE HALF	VULGAR FRACTION ONE HALF
00BE	FRACTION THREE QUARTERS	VULGAR FRACTION THREE QUARTERS
Also note in particular:		
FEFF	BYTE ORDER MARK	ZERO WIDTH NO-BREAK SPACE

The character U+FEFF retains the semantics of *byte order mark*, but it has been renamed to ZERO WIDTH NON-BREAK SPACE and given additional semantics as described *Section 2.4, Special Character and Non-Character Values*.

Character Semantics Changes

Joiners no longer can be used to request use of ligatures, nor to request that they not be used. The *byte order mark* gained additional semantics, as noted earlier.

D.3 Changes from Unicode 1.1 to Unicode 2.0

No characters have been removed or unified from the Unicode Standard, Version 1.1 to the Unicode Standard, Version 2.0.

Areas Redefined

Surrogates Area. In the Unicode Standard, Version 2.0, the range U+D800 → U+DFFF is newly defined as the Surrogates Area (see *Chapter 6, Character Block Descriptions*).

Hangul Syllables Area. The the Unicode Standard, Version 1.0 Hangul Area (U+3400 → U+3D2D) and the the Unicode Standard, Version 1.1 Hangul Supplementary Syllables A and B areas (U+3D2E → U+4DFF) containing 4,306 Korean Hangul syllables have been removed. In their stead in the Unicode Standard, Version 2.0, the Hangul Syllables Area (U+AC00 → U+D7A3) containing 11,172 Korean Hangul syllables has been added.

Characters Moved

The redefinition of the Hangul Syllables Area resulted in the movement of previously-encoded Hangul Syllables to the new area. No other characters have been moved in the transition from the Unicode Standard, Version 1.1 to the Unicode Standard, Version 2.0.

New Characters Added

The redefinition of the Hangul Syllables Area resulted in the addition of Hangul Syllables to the new area to complete the modern 11,172 Korean Hangul Syllable set.

In the Unicode Standard, Version 2.0, an additional 201 items also included in amendments to ISO 10646 were annexed into code positions that had been unassigned in the Unicode Standard, Version 1.1. These additions are listed below by block.

In the Hebrew block, 31 additional cantillation marks:

U+0591 → U+05AF
U+05C4 HEBREW MARK UPPER DOT

The whole Tibetan block, comprising 168 additions:

U+0F00 → U+0FBF

In the Latin Extended Additional block, 1 additional precomposed character:

U+1E9B LATIN SMALL LETTER LONG S WITH DOT ABOVE

In block Currency Symbols, 1 addition from Vietnam:

U+20AB DONG SIGN

Character Name Changes

LIGATURE was changed to LETTER in these 6 names:

Table D-7. 2.0 Name Changes

Code	Unicode 1.1	Unicode 2.0
00C6	LATIN CAPITAL LIGATURE AE	LATIN CAPITAL LETTER AE
00E6	LATIN SMALL LIGATURE AE	LATIN SMALL LETTER AE
01E2	LATIN CAPITAL LIGATURE AE WITH MACRON	LATIN CAPITAL LETTER AE WITH MACRON
01E3	LATIN SMALL LIGATURE AE WITH MACRON	LATIN SMALL LETTER AE WITH MACRON
01FC	LATIN CAPITAL LIGATURE AE WITH ACUTE	LATIN CAPITAL LETTER AE WITH ACUTE
01FD	LATIN SMALL LIGATURE AE WITH ACUTE	LATIN SMALL LETTER AE WITH ACUTE

ISO/IEC JTC1/SC2/WG2 and the Unicode Technical Committee have resolved that there shall be no further character name changes in the future.

Character Semantics Changes

The order of joiner characters used for Indic half-forms changed.

The Unicode Standard, Version 1.1 *Appendix G, Symmetric Swapping* became the normative property *mirrored*. See *Section 4.7, Mirrored*. See also ISO/IEC 10646, Annex C (Informative).