Chapter 4

Character Properties

> ***Disclaimer***
>
> The content of all character property tables has been verified as far as possible by the Unicode Consortium. However, the Unicode Consortium does not guarantee that the tables printed in this volume or on the CD-ROM are correct in every detail, and it is not responsible for errors that may occur either in the character property tables or in software that implements these tables. *The contents of all the tables in this chapter may be superseded or augmented by information on the Unicode Web site.*

This chapter describes the attributes of character properties defined by the Unicode Standard and gives mappings of characters to specific character properties. Full listings for all Unicode properties are provided in the *Unicode Character Database* on the CD-ROM.

While the Unicode Consortium strives to minimize changes to character property data, occasionally character properties must be updated. When this situation occurs, the relevant data files of the Unicode Character Database are revised. The revised data files are posted on the Unicode Web site as an update version of the standard.

Normative and Informative Properties

As specified in *Chapter 3, Conformance*, the Unicode Standard defines both normative and informative properties.

Normative Properties. *Normative* means that implementations that claim conformance to the Unicode Standard (at a particular version) and that make use of a particular property must follow the specifications of the standard for that property to be conformant. The term *normative* when applied to a character property does *not* mean that the value of the property will never change. Corrections and extensions to the standard in the future may require minor changes to normative values, even though the Unicode Technical Committee strives to minimize such changes.

Informative Properties. If a character property is only *informative*, a conformant implementation is free to use or change such values as it may require, while still remaining conformant to the standard. Particular implementations may choose to override the properties that are not normative. In that case, the implementer has the option of establishing a protocol to convey that information.

The normative character properties of the Unicode Standard are given in *Table 4-1*, and the informative character properties are given in *Table 4-2*. Also included in the tables are the locations of each property's description and of the list of characters with that property.

Table 4-1. Normative Character Properties

Property	Description	List
Case	Chapter 4	Chapter 14[a]
Combining Classes	Chapter 3	*Table 4-3*
Conjoining Jamo (1100–11FF)	Chapter 3	Chapter 14
Decomposition (Canonical and Compatibility)	Chapter 3	Chapter 14
Directionality	*Section 3.12*	CD-ROM
Jamo Short Name	Chapter 3	*Table 4-4*
Numeric Value	Chapter 4	*Table 4-6*
Private Use	Chapter 3	*Section 13.5*
Special Character Properties	Chapter 13	*Section 3.9*
Surrogate	Chapter 3	*Section 13.4*
Mirrored	Chapter 3	*Table 4-9*
Unicode Character Names	Chapter 14	Chapter 14

a. This property is listed by individual character where it is not obvious.

Table 4-2. Informative Character Properties

Property	Description	List
Case Mapping	Chapter 4	CD-ROM
Dashes	Chapter 6	*Table 6-2*
East Asian Width	*Section 10.3*	CD-ROM UTR #11
Letters (Alphabetic and Ideographic)	Chapter 4	CD-ROM
Line Breaking	Chapter 13	*Section 13.1* and *Section 13.2*
Mathematical Property	Chapter 4	*Section 4.9*
Spaces	Chapter 6	*Table 6-1*
Unicode 1.0 Names	Chapter 4	CD-ROM

Default Properties. Implementations need specific properties for all code points, including those that are unassigned. To meet this need, the Unicode standard assigns default properties to ranges of unassigned code points. For example, see *Table 3-8*.

Consistency of Properties. The Unicode Standard is the product of many compromises. It has to strike a balance between uniformity of treatment for similar characters and compatibility with existing practice for characters inherited from legacy encodings. Because of this balancing act, one can expect a certain number of anomalies in character properties. For example, some pairs of characters might have been treated as canonical equivalents but are left unequivalent for compatibility with legacy differences. This situation pertains to U+00B5 µ MICRO SIGN (cf. U+03BC μ GREEK SMALL LETTER MU) as well as to certain Korean jamo.

In addition, some characters might have had properties differing in some ways from those assigned in this standard, but whose properties are left as is for compatibility with existing practice. This situation can be seen with the halfwidth voicing marks for Japanese (U+FF9E HALFWIDTH KATAKANA VOICED SOUND MARK and U+FF9F HALFWIDTH KATAKANA SEMI-VOICED SOUND MARK), which might have been better analyzed as spacing combining marks, and with the conjoining Hangul jamo, which might have been better analyzed as an initial base character, followed by formally combining medial and final characters. In the interest of efficiency and uniformity in algorithms, implementations may take advantage of such reanalyses of character properties, as long as the results they produce do not overtly conflict with those specified by the normative properties of this standard.

4.1 Case—Normative

Case is a normative property of characters in certain alphabets whereby characters are considered to be variants of a single letter. These variants, which may differ markedly in shape and size, are called the *uppercase* letter (also known as *capital* or *majuscule*) and the *lowercase* letter (also known as *small* or *minuscule*). The uppercase letter is generally larger than the lowercase letter.

Because of the inclusion of certain composite characters for compatibility, such as U+01F1 LATIN CAPITAL LETTER DZ, a third case, called *titlecase*, is used where the first character of a word must be capitalized. An example of such a character is U+01F2 LATIN CAPITAL LETTER D WITH SMALL LETTER Z. The three case forms are UPPERCASE, Titlecase, lowercase.

For those scripts that have case (Latin, Greek, Cyrillic, Armenian, and archaic Georgian), the case of a Unicode character can usually be obtained from the character's name. This statement is true for only these five scripts. Uppercase characters typically contain the word *capital* in their names. Lowercase characters typically contain the word *small*. The word *small* in the names of characters from scripts other than the five just listed has nothing to do with case. (Note that while the archaic Georgian script contained upper- and lowercase pairs, they are rarely used in modern Georgian. See *Section 7.5, Georgian*.)

Case Mappings. The lowercase letter default case mapping occurs between the small character and the capital character. The Unicode Standard case mapping tables, which are informative, are on the CD-ROM. Exceptions to the normal casing rules can be found in the data file SpecialCasing.txt. For more information on case mappings, see *Section 5.18, Case Mappings*, and Unicode Technical Report #21, "Case Mappings," on the CD-ROM or the up-to-date version on the Unicode Web site.

4.2 Combining Classes—Normative

Each combining character has a normative canonical *combining class*. This class is used with the canonical ordering algorithm to determine which combining characters interact typographically and to determine how the canonical ordering of sequences of combining characters takes place. Class *zero* combining characters act like base letters for the purpose of determining canonical order. Combining characters with non-zero classes participate in reordering for the purpose of determining the canonical form of sequences of characters. (See *Section 3.10, Canonical Ordering Behavior*, for a description of the algorithm.)

The list of combining characters and their canonical combining class appears in *Table 4-3*. Most combining characters are nonspacing. The spacing, class zero, combining characters are so noted.

Table 4-3. Combining Classes

Spacing and Nonspacing (Class 0)		
U+07A6	0	THAANA ABAFILI
U+07A7	0	THAANA AABAAFILI
U+07A8	0	THAANA IBIFILI
U+07A9	0	THAANA EEBEEFILI
U+07AA	0	THAANA UBUFILI
U+07AB	0	THAANA OOBOOFILI
U+07AC	0	THAANA EBEFILI
U+07AD	0	THAANA EYBEYFILI
U+07AE	0	THAANA OBOFILI
U+07AF	0	THAANA OABOAFILI

Table 4-3. Combining Classes (Continued)

U+07B0	0	THAANA SUKUN
U+0901	0	DEVANAGARI SIGN CANDRABINDU
U+0902	0	DEVANAGARI SIGN ANUSVARA
U+0903	0	DEVANAGARI SIGN VISARGA
U+093E	0	DEVANAGARI VOWEL SIGN AA
U+0940	0	DEVANAGARI VOWEL SIGN II
U+0941	0	DEVANAGARI VOWEL SIGN U
U+0942	0	DEVANAGARI VOWEL SIGN UU
U+0943	0	DEVANAGARI VOWEL SIGN VOCALIC R
U+0944	0	DEVANAGARI VOWEL SIGN VOCALIC RR
U+0945	0	DEVANAGARI VOWEL SIGN CANDRA E
U+0946	0	DEVANAGARI VOWEL SIGN SHORT E
U+0947	0	DEVANAGARI VOWEL SIGN E
U+0948	0	DEVANAGARI VOWEL SIGN AI
U+0949	0	DEVANAGARI VOWEL SIGN CANDRA O
U+094A	0	DEVANAGARI VOWEL SIGN SHORT O
U+094B	0	DEVANAGARI VOWEL SIGN O
U+094C	0	DEVANAGARI VOWEL SIGN AU
U+0962	0	DEVANAGARI VOWEL SIGN VOCALIC L
U+0963	0	DEVANAGARI VOWEL SIGN VOCALIC LL
U+0981	0	BENGALI SIGN CANDRABINDU
U+0982	0	BENGALI SIGN ANUSVARA
U+0983	0	BENGALI SIGN VISARGA
U+09BE	0	BENGALI VOWEL SIGN AA
U+09C0	0	BENGALI VOWEL SIGN II
U+09C1	0	BENGALI VOWEL SIGN U
U+09C2	0	BENGALI VOWEL SIGN UU
U+09C3	0	BENGALI VOWEL SIGN VOCALIC R
U+09C4	0	BENGALI VOWEL SIGN VOCALIC RR
U+09D7	0	BENGALI AU LENGTH MARK
U+09E2	0	BENGALI VOWEL SIGN VOCALIC L
U+09E3	0	BENGALI VOWEL SIGN VOCALIC LL
U+0A02	0	GURMUKHI SIGN BINDI
U+0A3E	0	GURMUKHI VOWEL SIGN AA
U+0A40	0	GURMUKHI VOWEL SIGN II
U+0A41	0	GURMUKHI VOWEL SIGN U
U+0A42	0	GURMUKHI VOWEL SIGN UU
U+0A47	0	GURMUKHI VOWEL SIGN EE
U+0A48	0	GURMUKHI VOWEL SIGN AI
U+0A4B	0	GURMUKHI VOWEL SIGN OO
U+0A4C	0	GURMUKHI VOWEL SIGN AU
U+0A70	0	GURMUKHI TIPPI
U+0A71	0	GURMUKHI ADDAK
U+0A81	0	GUJARATI SIGN CANDRABINDU
U+0A82	0	GUJARATI SIGN ANUSVARA
U+0A83	0	GUJARATI SIGN VISARGA
U+0ABE	0	GUJARATI VOWEL SIGN AA
U+0AC0	0	GUJARATI VOWEL SIGN II
U+0AC1	0	GUJARATI VOWEL SIGN U
U+0AC2	0	GUJARATI VOWEL SIGN UU
U+0AC3	0	GUJARATI VOWEL SIGN VOCALIC R
U+0AC4	0	GUJARATI VOWEL SIGN VOCALIC RR
U+0AC5	0	GUJARATI VOWEL SIGN CANDRA E
U+0AC7	0	GUJARATI VOWEL SIGN E
U+0AC8	0	GUJARATI VOWEL SIGN AI
U+0AC9	0	GUJARATI VOWEL SIGN CANDRA O
U+0ACB	0	GUJARATI VOWEL SIGN O
U+0ACC	0	GUJARATI VOWEL SIGN AU
U+0B01	0	ORIYA SIGN CANDRABINDU
U+0B02	0	ORIYA SIGN ANUSVARA
U+0B03	0	ORIYA SIGN VISARGA
U+0B3E	0	ORIYA VOWEL SIGN AA

Table 4-3. Combining Classes (Continued)

U+0B3F	0	ORIYA VOWEL SIGN I
U+0B40	0	ORIYA VOWEL SIGN II
U+0B41	0	ORIYA VOWEL SIGN U
U+0B42	0	ORIYA VOWEL SIGN UU
U+0B43	0	ORIYA VOWEL SIGN VOCALIC R
U+0B56	0	ORIYA AI LENGTH MARK
U+0B57	0	ORIYA AU LENGTH MARK
U+0B82	0	TAMIL SIGN ANUSVARA
U+0B83	0	TAMIL SIGN VISARGA
U+0BBE	0	TAMIL VOWEL SIGN AA
U+0BBF	0	TAMIL VOWEL SIGN I
U+0BC0	0	TAMIL VOWEL SIGN II
U+0BC1	0	TAMIL VOWEL SIGN U
U+0BC2	0	TAMIL VOWEL SIGN UU
U+0BD7	0	TAMIL AU LENGTH MARK
U+0C01	0	TELUGU SIGN CANDRABINDU
U+0C02	0	TELUGU SIGN ANUSVARA
U+0C03	0	TELUGU SIGN VISARGA
U+0C3E	0	TELUGU VOWEL SIGN AA
U+0C3F	0	TELUGU VOWEL SIGN I
U+0C40	0	TELUGU VOWEL SIGN II
U+0C41	0	TELUGU VOWEL SIGN U
U+0C42	0	TELUGU VOWEL SIGN UU
U+0C43	0	TELUGU VOWEL SIGN VOCALIC R
U+0C44	0	TELUGU VOWEL SIGN VOCALIC RR
U+0C46	0	TELUGU VOWEL SIGN E
U+0C47	0	TELUGU VOWEL SIGN EE
U+0C48	0	TELUGU VOWEL SIGN AI
U+0C4A	0	TELUGU VOWEL SIGN O
U+0C4B	0	TELUGU VOWEL SIGN OO
U+0C4C	0	TELUGU VOWEL SIGN AU
U+0C82	0	KANNADA SIGN ANUSVARA
U+0C83	0	KANNADA SIGN VISARGA
U+0CBE	0	KANNADA VOWEL SIGN AA
U+0CBF	0	KANNADA VOWEL SIGN I
U+0CC1	0	KANNADA VOWEL SIGN U
U+0CC2	0	KANNADA VOWEL SIGN UU
U+0CC3	0	KANNADA VOWEL SIGN VOCALIC R
U+0CC4	0	KANNADA VOWEL SIGN VOCALIC RR
U+0CC6	0	KANNADA VOWEL SIGN E
U+0CCC	0	KANNADA VOWEL SIGN AU
U+0CD5	0	KANNADA LENGTH MARK
U+0CD6	0	KANNADA AI LENGTH MARK
U+0D02	0	MALAYALAM SIGN ANUSVARA
U+0D03	0	MALAYALAM SIGN VISARGA
U+0D3E	0	MALAYALAM VOWEL SIGN AA
U+0D3F	0	MALAYALAM VOWEL SIGN I
U+0D40	0	MALAYALAM VOWEL SIGN II
U+0D41	0	MALAYALAM VOWEL SIGN U
U+0D42	0	MALAYALAM VOWEL SIGN UU
U+0D43	0	MALAYALAM VOWEL SIGN VOCALIC R
U+0D57	0	MALAYALAM AU LENGTH MARK
U+0D82	0	SINHALA SIGN ANUSVARAYA
U+0D83	0	SINHALA SIGN VISARGAYA
U+0DCF	0	SINHALA VOWEL SIGN AELA-PILLA
U+0DD0	0	SINHALA VOWEL SIGN KETTI AEDA-PILLA
U+0DD1	0	SINHALA VOWEL SIGN DIGA AEDA-PILLA
U+0DD2	0	SINHALA VOWEL SIGN KETTI IS-PILLA
U+0DD3	0	SINHALA VOWEL SIGN DIGA IS-PILLA
U+0DD4	0	SINHALA VOWEL SIGN KETTI PAA-PILLA
U+0DD6	0	SINHALA VOWEL SIGN DIGA PAA-PILLA
U+0DD8	0	SINHALA VOWEL SIGN GAETTA-PILLA

Table 4-3. Combining Classes (Continued)

U+0DDF	0	SINHALA VOWEL SIGN GAYANUKITTA
U+0DF2	0	SINHALA VOWEL SIGN DIGA GAETTA-PILLA
U+0DF3	0	SINHALA VOWEL SIGN DIGA GAYANUKITTA
U+0E31	0	THAI CHARACTER MAI HAN-AKAT
U+0E34	0	THAI CHARACTER SARA I
U+0E35	0	THAI CHARACTER SARA II
U+0E36	0	THAI CHARACTER SARA UE
U+0E37	0	THAI CHARACTER SARA UEE
U+0E47	0	THAI CHARACTER MAITAIKHU
U+0E4C	0	THAI CHARACTER THANTHAKHAT
U+0E4D	0	THAI CHARACTER NIKHAHIT
U+0E4E	0	THAI CHARACTER YAMAKKAN
U+0EB1	0	LAO VOWEL SIGN MAI KAN
U+0EB4	0	LAO VOWEL SIGN I
U+0EB5	0	LAO VOWEL SIGN II
U+0EB6	0	LAO VOWEL SIGN Y
U+0EB7	0	LAO VOWEL SIGN YY
U+0EBB	0	LAO VOWEL SIGN MAI KON
U+0EBC	0	LAO SEMIVOWEL SIGN LO
U+0ECC	0	LAO CANCELLATION MARK
U+0ECD	0	LAO NIGGAHITA
U+0F3E	0	TIBETAN SIGN YAR TSHES
U+0F3F	0	TIBETAN SIGN MAR TSHES
U+0F73	0	TIBETAN VOWEL SIGN II
U+0F75	0	TIBETAN VOWEL SIGN UU
U+0F76	0	TIBETAN VOWEL SIGN VOCALIC R
U+0F77	0	TIBETAN VOWEL SIGN VOCALIC RR
U+0F78	0	TIBETAN VOWEL SIGN VOCALIC L
U+0F79	0	TIBETAN VOWEL SIGN VOCALIC LL
U+0F7E	0	TIBETAN SIGN RJES SU NGA RO
U+0F7F	0	TIBETAN SIGN RNAM BCAD
U+0F81	0	TIBETAN VOWEL SIGN REVERSED II
U+102C	0	MYANMAR VOWEL SIGN AA
U+102D	0	MYANMAR VOWEL SIGN I
U+102E	0	MYANMAR VOWEL SIGN II
U+102F	0	MYANMAR VOWEL SIGN U
U+1030	0	MYANMAR VOWEL SIGN UU
U+1032	0	MYANMAR VOWEL SIGN AI
U+1036	0	MYANMAR SIGN ANUSVARA
U+1038	0	MYANMAR SIGN VISARGA
U+1056	0	MYANMAR VOWEL SIGN VOCALIC R
U+1057	0	MYANMAR VOWEL SIGN VOCALIC RR
U+1058	0	MYANMAR VOWEL SIGN VOCALIC L
U+1059	0	MYANMAR VOWEL SIGN VOCALIC LL
U+17B4	0	KHMER VOWEL INHERENT AQ
U+17B5	0	KHMER VOWEL INHERENT AA
U+17B6	0	KHMER VOWEL SIGN AA
U+17B7	0	KHMER VOWEL SIGN I
U+17B8	0	KHMER VOWEL SIGN II
U+17B9	0	KHMER VOWEL SIGN Y
U+17BA	0	KHMER VOWEL SIGN YY
U+17BB	0	KHMER VOWEL SIGN U
U+17BC	0	KHMER VOWEL SIGN UU
U+17BD	0	KHMER VOWEL SIGN UA
U+17C6	0	KHMER SIGN NIKAHIT
U+17C7	0	KHMER SIGN REAHMUK
U+17C8	0	KHMER SIGN YUUKALEAPINTU
U+17C9	0	KHMER SIGN MUUSIKATOAN
U+17CA	0	KHMER SIGN TRIISAP
U+17CB	0	KHMER SIGN BANTOC
U+17CC	0	KHMER SIGN ROBAT
U+17CD	0	KHMER SIGN TOANDAKHIAT

Table 4-3. Combining Classes (Continued)

U+17CE	0	KHMER SIGN KAKABAT
U+17CF	0	KHMER SIGN AHSDA
U+17D0	0	KHMER SIGN SAMYOK SANNYA
U+17D1	0	KHMER SIGN VIRIAM
U+17D3	0	KHMER SIGN BATHAMASAT
Split		**(Class 0)**
U+09CB	0	BENGALI VOWEL SIGN O
U+09CC	0	BENGALI VOWEL SIGN AU
U+0B48	0	ORIYA VOWEL SIGN AI
U+0B4B	0	ORIYA VOWEL SIGN O
U+0B4C	0	ORIYA VOWEL SIGN AU
U+0BCA	0	TAMIL VOWEL SIGN O
U+0BCB	0	TAMIL VOWEL SIGN OO
U+0BCC	0	TAMIL VOWEL SIGN AU
U+0CC0	0	KANNADA VOWEL SIGN II
U+0CC7	0	KANNADA VOWEL SIGN EE
U+0CC8	0	KANNADA VOWEL SIGN AI
U+0CCA	0	KANNADA VOWEL SIGN O
U+0CCB	0	KANNADA VOWEL SIGN OO
U+0D4A	0	MALAYALAM VOWEL SIGN O
U+0D4B	0	MALAYALAM VOWEL SIGN OO
U+0D4C	0	MALAYALAM VOWEL SIGN AU
U+0DDC	0	SINHALA VOWEL SIGN KOMBUVA HAA AELA-PILLA
U+0DDD	0	SINHALA VOWEL SIGN KOMBUVA HAA DIGA AELA-PILLA
U+0DDE	0	SINHALA VOWEL SIGN KOMBUVA HAA GAYANUKITTA
U+17BF	0	KHMER VOWEL SIGN YA
U+17C0	0	KHMER VOWEL SIGN IE
U+17C4	0	KHMER VOWEL SIGN OO
U+17C5	0	KHMER VOWEL SIGN AU
Reordrant		**(Class 0)**
U+093F	0	DEVANAGARI VOWEL SIGN I
U+09BF	0	BENGALI VOWEL SIGN I
U+09C7	0	BENGALI VOWEL SIGN E
U+09C8	0	BENGALI VOWEL SIGN AI
U+0A3F	0	GURMUKHI VOWEL SIGN I
U+0ABF	0	GUJARATI VOWEL SIGN I
U+0B47	0	ORIYA VOWEL SIGN E
U+0BC6	0	TAMIL VOWEL SIGN E
U+0BC7	0	TAMIL VOWEL SIGN EE
U+0BC8	0	TAMIL VOWEL SIGN AI
U+0D46	0	MALAYALAM VOWEL SIGN E
U+0D47	0	MALAYALAM VOWEL SIGN EE
U+0D48	0	MALAYALAM VOWEL SIGN AI
U+0DD9	0	SINHALA VOWEL SIGN KOMBUVA
U+0DDA	0	SINHALA VOWEL SIGN DIGA KOMBUVA
U+0DDB	0	SINHALA VOWEL SIGN KOMBU DEKA
U+1031	0	MYANMAR VOWEL SIGN E
U+17BE	0	KHMER VOWEL SIGN OE
U+17C1	0	KHMER VOWEL SIGN E
U+17C2	0	KHMER VOWEL SIGN AE
U+17C3	0	KHMER VOWEL SIGN AI
Tibetan Subjoined Letters		**(Class 0)**
U+0F90	0	TIBETAN SUBJOINED LETTER KA
U+0F91	0	TIBETAN SUBJOINED LETTER KHA
U+0F92	0	TIBETAN SUBJOINED LETTER GA
U+0F93	0	TIBETAN SUBJOINED LETTER GHA
U+0F94	0	TIBETAN SUBJOINED LETTER NGA
U+0F95	0	TIBETAN SUBJOINED LETTER CA
U+0F96	0	TIBETAN SUBJOINED LETTER CHA
U+0F97	0	TIBETAN SUBJOINED LETTER JA

Table 4-3. Combining Classes (Continued)

U+0F99	0	TIBETAN SUBJOINED LETTER NYA
U+0F9A	0	TIBETAN SUBJOINED LETTER TTA
U+0F9B	0	TIBETAN SUBJOINED LETTER TTHA
U+0F9C	0	TIBETAN SUBJOINED LETTER DDA
U+0F9D	0	TIBETAN SUBJOINED LETTER DDHA
U+0F9E	0	TIBETAN SUBJOINED LETTER NNA
U+0F9F	0	TIBETAN SUBJOINED LETTER TA
U+0FA0	0	TIBETAN SUBJOINED LETTER THA
U+0FA1	0	TIBETAN SUBJOINED LETTER DA
U+0FA2	0	TIBETAN SUBJOINED LETTER DHA
U+0FA3	0	TIBETAN SUBJOINED LETTER NA
U+0FA4	0	TIBETAN SUBJOINED LETTER PA
U+0FA5	0	TIBETAN SUBJOINED LETTER PHA
U+0FA6	0	TIBETAN SUBJOINED LETTER BA
U+0FA7	0	TIBETAN SUBJOINED LETTER BHA
U+0FA8	0	TIBETAN SUBJOINED LETTER MA
U+0FA9	0	TIBETAN SUBJOINED LETTER TSA
U+0FAA	0	TIBETAN SUBJOINED LETTER TSHA
U+0FAB	0	TIBETAN SUBJOINED LETTER DZA
U+0FAC	0	TIBETAN SUBJOINED LETTER DZHA
U+0FAD	0	TIBETAN SUBJOINED LETTER WA
U+0FAE	0	TIBETAN SUBJOINED LETTER ZHA
U+0FAF	0	TIBETAN SUBJOINED LETTER ZA
U+0FB0	0	TIBETAN SUBJOINED LETTER -A
U+0FB1	0	TIBETAN SUBJOINED LETTER YA
U+0FB2	0	TIBETAN SUBJOINED LETTER RA
U+0FB3	0	TIBETAN SUBJOINED LETTER LA
U+0FB4	0	TIBETAN SUBJOINED LETTER SHA
U+0FB5	0	TIBETAN SUBJOINED LETTER SSA
U+0FB6	0	TIBETAN SUBJOINED LETTER SA
U+0FB7	0	TIBETAN SUBJOINED LETTER HA
U+0FB8	0	TIBETAN SUBJOINED LETTER A
U+0FB9	0	TIBETAN SUBJOINED LETTER KSSA
U+0FBA	0	TIBETAN SUBJOINED LETTER FIXED-FORM WA
U+0FBB	0	TIBETAN SUBJOINED LETTER FIXED-FORM YA
U+0FBC	0	TIBETAN SUBJOINED LETTER FIXED-FORM RA
Enclosing		**(Class 0)**
U+0488	0	CYRILLIC HUNDRED THOUSANDS SIGN
U+0489	0	CYRILLIC MILLIONS SIGN
U+06DD	0	ARABIC END OF AYAH
U+06DE	0	ARABIC START OF RUB EL HIZB
U+20DD	0	COMBINING ENCLOSING CIRCLE
U+20DE	0	COMBINING ENCLOSING SQUARE
U+20DF	0	COMBINING ENCLOSING DIAMOND
U+20E0	0	COMBINING ENCLOSING CIRCLE BACKSLASH
U+20E2	0	COMBINING ENCLOSING SCREEN
U+20E3	0	COMBINING ENCLOSING KEYCAP
Overlays/Interior		**(Class 1)**
U+0334	1	COMBINING TILDE OVERLAY
U+0335	1	COMBINING SHORT STROKE OVERLAY
U+0336	1	COMBINING LONG STROKE OVERLAY
U+0337	1	COMBINING SHORT SOLIDUS OVERLAY
U+0338	1	COMBINING LONG SOLIDUS OVERLAY
U+20D2	1	COMBINING LONG VERTICAL LINE OVERLAY
U+20D3	1	COMBINING SHORT VERTICAL LINE OVERLAY
U+20D8	1	COMBINING RING OVERLAY
U+20D9	1	COMBINING CLOCKWISE RING OVERLAY
U+20DA	1	COMBINING ANTICLOCKWISE RING OVERLAY
Nuktas		**(Class 7)**
U+093C	7	DEVANAGARI SIGN NUKTA
U+09BC	7	BENGALI SIGN NUKTA

Table 4-3. Combining Classes (Continued)

U+0A3C	7	GURMUKHI SIGN NUKTA
U+0ABC	7	GUJARATI SIGN NUKTA
U+0B3C	7	ORIYA SIGN NUKTA
U+1037	7	MYANMAR SIGN DOT BELOW
Kana Voicing Marks		**(Class 8)**
U+3099	8	COMBINING KATAKANA-HIRAGANA VOICED SOUND MARK
U+309A	8	COMBINING KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK
Viramas		**(Class 9)**
U+094D	9	DEVANAGARI SIGN VIRAMA
U+09CD	9	BENGALI SIGN VIRAMA
U+0A4D	9	GURMUKHI SIGN VIRAMA
U+0ACD	9	GUJARATI SIGN VIRAMA
U+0B4D	9	ORIYA SIGN VIRAMA
U+0BCD	9	TAMIL SIGN VIRAMA
U+0C4D	9	TELUGU SIGN VIRAMA
U+0CCD	9	KANNADA SIGN VIRAMA
U+0D4D	9	MALAYALAM SIGN VIRAMA
U+0DCA	9	SINHALA SIGN AL-LAKUNA
U+0E3A	9	THAI CHARACTER PHINTHU
U+0F84	9	TIBETAN MARK HALANTA
U+1039	9	MYANMAR SIGN VIRAMA
U+17D2	9	KHMER SIGN COENG
Fixed Position Classes		**(Classes 10..199)**
U+05B0	10	HEBREW POINT SHEVA
U+05B1	11	HEBREW POINT HATAF SEGOL
U+05B2	12	HEBREW POINT HATAF PATAH
U+05B3	13	HEBREW POINT HATAF QAMATS
U+05B4	14	HEBREW POINT HIRIQ
U+05B5	15	HEBREW POINT TSERE
U+05B6	16	HEBREW POINT SEGOL
U+05B7	17	HEBREW POINT PATAH
U+05B8	18	HEBREW POINT QAMATS
U+05B9	19	HEBREW POINT HOLAM
U+05BB	20	HEBREW POINT QUBUTS
U+05BC	21	HEBREW POINT DAGESH OR MAPIQ
U+05BD	22	HEBREW POINT METEG
U+05BF	23	HEBREW POINT RAFE
U+05C1	24	HEBREW POINT SHIN DOT
U+05C2	25	HEBREW POINT SIN DOT
U+FB1E	26	HEBREW POINT JUDEO-SPANISH VARIKA
U+064B	27	ARABIC FATHATAN
U+064C	28	ARABIC DAMMATAN
U+064D	29	ARABIC KASRATAN
U+064E	30	ARABIC FATHA
U+064F	31	ARABIC DAMMA
U+0650	32	ARABIC KASRA
U+0651	33	ARABIC SHADDA
U+0652	34	ARABIC SUKUN
U+0670	35	ARABIC LETTER SUPERSCRIPT ALEF
U+0711	36	SYRIAC LETTER SUPERSCRIPT ALAPH
U+0C55	84	TELUGU LENGTH MARK
U+0C56	91	TELUGU AI LENGTH MARK
U+0E38	103	THAI CHARACTER SARA U
U+0E39	103	THAI CHARACTER SARA UU
U+0E48	107	THAI CHARACTER MAI EK
U+0E49	107	THAI CHARACTER MAI THO
U+0E4A	107	THAI CHARACTER MAI TRI
U+0E4B	107	THAI CHARACTER MAI CHATTAWA
U+0EB8	118	LAO VOWEL SIGN U
U+0EB9	118	LAO VOWEL SIGN UU

Table 4-3. Combining Classes (Continued)

U+0EC8	122	LAO TONE MAI EK
U+0EC9	122	LAO TONE MAI THO
U+0ECA	122	LAO TONE MAI TI
U+0ECB	122	LAO TONE MAI CATAWA
U+0F71	129	TIBETAN VOWEL SIGN AA
U+0F72	130	TIBETAN VOWEL SIGN I
U+0F7A	130	TIBETAN VOWEL SIGN E
U+0F7B	130	TIBETAN VOWEL SIGN EE
U+0F7C	130	TIBETAN VOWEL SIGN O
U+0F7D	130	TIBETAN VOWEL SIGN OO
U+0F80	130	TIBETAN VOWEL SIGN REVERSED I
U+0F74	132	TIBETAN VOWEL SIGN U
Below Left Attached		**(Class 200)**
Below Attached		**(Class 202)**
U+0321	202	COMBINING PALATALIZED HOOK BELOW
U+0322	202	COMBINING RETROFLEX HOOK BELOW
U+0327	202	COMBINING CEDILLA
U+0328	202	COMBINING OGONEK
Below Right Attached		**(Class 204)**
Left Attached Reordrant		**(Class 208)**
Right Attached		**(Class 210)**
Above Left Attached		**(Class 212)**
Above Attached		**(Class 214)**
Above Right Attached		**(Class 216)**
U+031B	216	COMBINING HORN
U+0F39	216	TIBETAN MARK TSA -PHRU
Below Left		**(Class 218)**
U+302A	218	IDEOGRAPHIC LEVEL TONE MARK
Below		**(Class 220)**
U+0316	220	COMBINING GRAVE ACCENT BELOW
U+0317	220	COMBINING ACUTE ACCENT BELOW
U+0318	220	COMBINING LEFT TACK BELOW
U+0319	220	COMBINING RIGHT TACK BELOW
U+031C	220	COMBINING LEFT HALF RING BELOW
U+031D	220	COMBINING UP TACK BELOW
U+031E	220	COMBINING DOWN TACK BELOW
U+031F	220	COMBINING PLUS SIGN BELOW
U+0320	220	COMBINING MINUS SIGN BELOW
U+0323	220	COMBINING DOT BELOW
U+0324	220	COMBINING DIAERESIS BELOW
U+0325	220	COMBINING RING BELOW
U+0326	220	COMBINING COMMA BELOW
U+0329	220	COMBINING VERTICAL LINE BELOW
U+032A	220	COMBINING BRIDGE BELOW
U+032B	220	COMBINING INVERTED DOUBLE ARCH BELOW
U+032C	220	COMBINING CARON BELOW
U+032D	220	COMBINING CIRCUMFLEX ACCENT BELOW
U+032E	220	COMBINING BREVE BELOW
U+032F	220	COMBINING INVERTED BREVE BELOW
U+0330	220	COMBINING TILDE BELOW
U+0331	220	COMBINING MACRON BELOW
U+0332	220	COMBINING LOW LINE
U+0333	220	COMBINING DOUBLE LOW LINE
U+0339	220	COMBINING RIGHT HALF RING BELOW
U+033A	220	COMBINING INVERTED BRIDGE BELOW
U+033B	220	COMBINING SQUARE BELOW
U+033C	220	COMBINING SEAGULL BELOW

Table 4-3. Combining Classes (Continued)

U+0347	220	COMBINING EQUALS SIGN BELOW
U+0348	220	COMBINING DOUBLE VERTICAL LINE BELOW
U+0349	220	COMBINING LEFT ANGLE BELOW
U+034D	220	COMBINING LEFT RIGHT ARROW BELOW
U+034E	220	COMBINING UPWARDS ARROW BELOW
U+0591	220	HEBREW ACCENT ETNAHTA
U+0596	220	HEBREW ACCENT TIPEHA
U+059B	220	HEBREW ACCENT TEVIR
U+05A3	220	HEBREW ACCENT MUNAH
U+05A4	220	HEBREW ACCENT MAHAPAKH
U+05A5	220	HEBREW ACCENT MERKHA
U+05A6	220	HEBREW ACCENT MERKHA KEFULA
U+05A7	220	HEBREW ACCENT DARGA
U+05AA	220	HEBREW ACCENT YERAH BEN YOMO
U+0655	220	ARABIC HAMZA BELOW
U+06E3	220	ARABIC SMALL LOW SEEN
U+06EA	220	ARABIC EMPTY CENTRE LOW STOP
U+06ED	220	ARABIC SMALL LOW MEEM
U+0731	220	SYRIAC PTHAHA BELOW
U+0734	220	SYRIAC ZQAPHA BELOW
U+0737	220	SYRIAC RBASA BELOW
U+0738	220	SYRIAC DOTTED ZLAMA HORIZONTAL
U+0739	220	SYRIAC DOTTED ZLAMA ANGULAR
U+073B	220	SYRIAC HBASA BELOW
U+073C	220	SYRIAC HBASA/ESASA DOTTED
U+073E	220	SYRIAC ESASA BELOW
U+0742	220	SYRIAC RUKKAKHA
U+0744	220	SYRIAC TWO VERTICAL DOTS BELOW
U+0746	220	SYRIAC THREE DOTS BELOW
U+0748	220	SYRIAC OBLIQUE LINE BELOW
U+0952	220	DEVANAGARI STRESS SIGN ANUDATTA
U+0F18	220	TIBETAN ASTROLOGICAL SIGN -KHYUD PA
U+0F19	220	TIBETAN ASTROLOGICAL SIGN SDONG TSHUGS
U+0F35	220	TIBETAN MARK NGAS BZUNG NYI ZLA
U+0F37	220	TIBETAN MARK NGAS BZUNG SGOR RTAGS
U+0FC6	220	TIBETAN SYMBOL PADMA GDAN
Below Right		**(Class 222)**
U+059A	222	HEBREW ACCENT YETIV
U+05AD	222	HEBREW ACCENT DEHI
U+302D	222	IDEOGRAPHIC ENTERING TONE MARK
Left		**(Class 224)**
U+302E	224	HANGUL SINGLE DOT TONE MARK
U+302F	224	HANGUL DOUBLE DOT TONE MARK
Right		**(Class 226)**
Above Left		**(Class 228)**
U+05AE	228	HEBREW ACCENT ZINOR
U+18A9	228	MONGOLIAN LETTER AG DAGALGA
U+302B	228	IDEOGRAPHIC RISING TONE MARK
Above		**(Class 230)**
U+0300	230	COMBINING GRAVE ACCENT
U+0301	230	COMBINING ACUTE ACCENT
U+0302	230	COMBINING CIRCUMFLEX ACCENT
U+0303	230	COMBINING TILDE
U+0304	230	COMBINING MACRON
U+0305	230	COMBINING OVERLINE
U+0306	230	COMBINING BREVE
U+0307	230	COMBINING DOT ABOVE
U+0308	230	COMBINING DIAERESIS
U+0309	230	COMBINING HOOK ABOVE
U+030A	230	COMBINING RING ABOVE

Table 4-3. Combining Classes (Continued)

U+030B	230	COMBINING DOUBLE ACUTE ACCENT
U+030C	230	COMBINING CARON
U+030D	230	COMBINING VERTICAL LINE ABOVE
U+030E	230	COMBINING DOUBLE VERTICAL LINE ABOVE
U+030F	230	COMBINING DOUBLE GRAVE ACCENT
U+0310	230	COMBINING CANDRABINDU
U+0311	230	COMBINING INVERTED BREVE
U+0312	230	COMBINING TURNED COMMA ABOVE
U+0313	230	COMBINING COMMA ABOVE
U+0314	230	COMBINING REVERSED COMMA ABOVE
U+033D	230	COMBINING X ABOVE
U+033E	230	COMBINING VERTICAL TILDE
U+033F	230	COMBINING DOUBLE OVERLINE
U+0340	230	COMBINING GRAVE TONE MARK
U+0341	230	COMBINING ACUTE TONE MARK
U+0342	230	COMBINING GREEK PERISPOMENI
U+0343	230	COMBINING GREEK KORONIS
U+0344	230	COMBINING GREEK DIALYTIKA TONOS
U+0346	230	COMBINING BRIDGE ABOVE
U+034A	230	COMBINING NOT TILDE ABOVE
U+034B	230	COMBINING HOMOTHETIC ABOVE
U+034C	230	COMBINING ALMOST EQUAL TO ABOVE
U+0483	230	COMBINING CYRILLIC TITLO
U+0484	230	COMBINING CYRILLIC PALATALIZATION
U+0485	230	COMBINING CYRILLIC DASIA PNEUMATA
U+0486	230	COMBINING CYRILLIC PSILI PNEUMATA
U+0592	230	HEBREW ACCENT SEGOL
U+0593	230	HEBREW ACCENT SHALSHELET
U+0594	230	HEBREW ACCENT ZAQEF QATAN
U+0595	230	HEBREW ACCENT ZAQEF GADOL
U+0597	230	HEBREW ACCENT REVIA
U+0598	230	HEBREW ACCENT ZARQA
U+0599	230	HEBREW ACCENT PASHTA
U+059C	230	HEBREW ACCENT GERESH
U+059D	230	HEBREW ACCENT GERESH MUQDAM
U+059E	230	HEBREW ACCENT GERSHAYIM
U+059F	230	HEBREW ACCENT QARNEY PARA
U+05A0	230	HEBREW ACCENT TELISHA GEDOLA
U+05A1	230	HEBREW ACCENT PAZER
U+05A8	230	HEBREW ACCENT QADMA
U+05A9	230	HEBREW ACCENT TELISHA QETANA
U+05AB	230	HEBREW ACCENT OLE
U+05AC	230	HEBREW ACCENT ILUY
U+05AF	230	HEBREW MARK MASORA CIRCLE
U+05C4	230	HEBREW MARK UPPER DOT
U+0653	230	ARABIC MADDAH ABOVE
U+0654	230	ARABIC HAMZA ABOVE
U+06D6	230	ARABIC SMALL HIGH LIGATURE SAD WITH LAM WITH ALEF MAKSURA
U+06D7	230	ARABIC SMALL HIGH LIGATURE QAF WITH LAM WITH ALEF MAKSURA
U+06D8	230	ARABIC SMALL HIGH MEEM INITIAL FORM
U+06D9	230	ARABIC SMALL HIGH LAM ALEF
U+06DA	230	ARABIC SMALL HIGH JEEM
U+06DB	230	ARABIC SMALL HIGH THREE DOTS
U+06DC	230	ARABIC SMALL HIGH SEEN
U+06DF	230	ARABIC SMALL HIGH ROUNDED ZERO
U+06E0	230	ARABIC SMALL HIGH UPRIGHT RECTANGULAR ZERO
U+06E1	230	ARABIC SMALL HIGH DOTLESS HEAD OF KHAH
U+06E2	230	ARABIC SMALL HIGH MEEM ISOLATED FORM
U+06E4	230	ARABIC SMALL HIGH MADDA
U+06E7	230	ARABIC SMALL HIGH YEH

Table 4-3. Combining Classes (Continued)

U+06E8	230	ARABIC SMALL HIGH NOON
U+06EB	230	ARABIC EMPTY CENTRE HIGH STOP
U+06EC	230	ARABIC ROUNDED HIGH STOP WITH FILLED CENTRE
U+0730	230	SYRIAC PTHAHA ABOVE
U+0732	230	SYRIAC PTHAHA DOTTED
U+0733	230	SYRIAC ZQAPHA ABOVE
U+0735	230	SYRIAC ZQAPHA DOTTED
U+0736	230	SYRIAC RBASA ABOVE
U+073A	230	SYRIAC HBASA ABOVE
U+073D	230	SYRIAC ESASA ABOVE
U+073F	230	SYRIAC RWAHA
U+0740	230	SYRIAC FEMININE DOT
U+0741	230	SYRIAC QUSHSHAYA
U+0743	230	SYRIAC TWO VERTICAL DOTS ABOVE
U+0745	230	SYRIAC THREE DOTS ABOVE
U+0747	230	SYRIAC OBLIQUE LINE ABOVE
U+0749	230	SYRIAC MUSIC
U+074A	230	SYRIAC BARREKH
U+0951	230	DEVANAGARI STRESS SIGN UDATTA
U+0953	230	DEVANAGARI GRAVE ACCENT
U+0954	230	DEVANAGARI ACUTE ACCENT
U+0F82	230	TIBETAN SIGN NYI ZLA NAA DA
U+0F83	230	TIBETAN SIGN SNA LDAN
U+0F86	230	TIBETAN SIGN LCI RTAGS
U+0F87	230	TIBETAN SIGN YANG RTAGS
U+20D0	230	COMBINING LEFT HARPOON ABOVE
U+20D1	230	COMBINING RIGHT HARPOON ABOVE
U+20D4	230	COMBINING ANTICLOCKWISE ARROW ABOVE
U+20D5	230	COMBINING CLOCKWISE ARROW ABOVE
U+20D6	230	COMBINING LEFT ARROW ABOVE
U+20D7	230	COMBINING RIGHT ARROW ABOVE
U+20DB	230	COMBINING THREE DOTS ABOVE
U+20DC	230	COMBINING FOUR DOTS ABOVE
U+20E1	230	COMBINING LEFT RIGHT ARROW ABOVE
U+FE20	230	COMBINING LIGATURE LEFT HALF
U+FE21	230	COMBINING LIGATURE RIGHT HALF
U+FE22	230	COMBINING DOUBLE TILDE LEFT HALF
U+FE23	230	COMBINING DOUBLE TILDE RIGHT HALF
Above Right		**(Class 232)**
U+0315	232	COMBINING COMMA ABOVE RIGHT
U+031A	232	COMBINING LEFT ANGLE ABOVE
U+302C	232	IDEOGRAPHIC DEPARTING TONE MARK
Double Below		**(Class 233)**
U+0362	233	COMBINING DOUBLE RIGHTWARDS ARROW BELOW
Double Above		**(Class 234)**
U+0360	234	COMBINING DOUBLE TILDE
U+0361	234	COMBINING DOUBLE INVERTED BREVE
Iota Subscript		**(Class 240)**
U+0345	240	COMBINING GREEK YPOGEGRAMMENI

4.3 Directionality—Normative

Directional behavior is interpreted according to the Unicode bidirectional algorithm (see *Section 3.12, Bidirectional Behavior*). For this purpose, all characters of the Unicode Standard possess a normative *directional* type. The directional types left-to-right and right-to-left are called *strong types*, and characters of these types are called strong directional characters. Left-to-right types include most alphabetic and syllabic characters, as well as all Han

ideographic characters. Right-to-left types include Arabic, Hebrew, Syriac, and Thaana, and most punctuation specific to those scripts. In addition, the Unicode bidirectional algorithm also uses *weak types* and *neutrals*.

For the directional types of Unicode characters, see the Unicode Character Database on the CD-ROM.

4.4 Jamo Short Names—Normative

The *jamo short name* is a normative property of the Unicode conjoining Hangul jamo characters. These short names, which are listed in *Table 4-4*, are used to determine the character names that are derived when decomposing Hangul syllables into their decomposition sequence.

Table 4-4. Jamo Short Names

Value	Short Name	Glyph	Unicode Name
U+1100	G	ᄀ	HANGUL CHOSEONG KIYEOK
U+1101	GG	ᄁ	HANGUL CHOSEONG SSANGKIYEOK
U+1102	N	ᄂ	HANGUL CHOSEONG NIEUN
U+1103	D	ᄃ	HANGUL CHOSEONG TIKEUT
U+1104	DD	ᄄ	HANGUL CHOSEONG SSANGTIKEUT
U+1105	R	ᄅ	HANGUL CHOSEONG RIEUL
U+1106	M	ᄆ	HANGUL CHOSEONG MIEUM
U+1107	B	ᄇ	HANGUL CHOSEONG PIEUP
U+1108	BB	ᄈ	HANGUL CHOSEONG SSANGPIEUP
U+1109	S	ᄉ	HANGUL CHOSEONG SIOS
U+110A	SS	ᄊ	HANGUL CHOSEONG SSANGSIOS
U+110B		ᄋ	HANGUL CHOSEONG IEUNG
U+110C	J	ᄌ	HANGUL CHOSEONG CIEUC
U+110D	JJ	ᄍ	HANGUL CHOSEONG SSANGCIEUC
U+110E	C	ᄎ	HANGUL CHOSEONG CHIEUCH
U+110F	K	ᄏ	HANGUL CHOSEONG KHIEUKH
U+1110	T	ᄐ	HANGUL CHOSEONG THIEUTH
U+1111	P	ᄑ	HANGUL CHOSEONG PHIEUPH
U+1112	H	ᄒ	HANGUL CHOSEONG HIEUH
U+1161	A	ᅡ	HANGUL JUNGSEONG A
U+1162	AE	ᅢ	HANGUL JUNGSEONG AE
U+1163	YA	ᅣ	HANGUL JUNGSEONG YA
U+1164	YAE	ᅤ	HANGUL JUNGSEONG YAE
U+1165	EO	ᅥ	HANGUL JUNGSEONG EO
U+1166	E	ᅦ	HANGUL JUNGSEONG E
U+1167	YEO	ᅧ	HANGUL JUNGSEONG YEO
U+1168	YE	ᅨ	HANGUL JUNGSEONG YE
U+1169	O	ᅩ	HANGUL JUNGSEONG O
U+116A	WA	ᅪ	HANGUL JUNGSEONG WA
U+116B	WAE	ᅫ	HANGUL JUNGSEONG WAE
U+116C	OE	ᅬ	HANGUL JUNGSEONG OE
U+116D	YO	ᅭ	HANGUL JUNGSEONG YO
U+116E	U	ᅮ	HANGUL JUNGSEONG U
U+116F	WEO	ᅯ	HANGUL JUNGSEONG WEO
U+1170	WE	ᅰ	HANGUL JUNGSEONG WE

Table 4-4. Jamo Short Names (Continued)

Value	Short Name	Glyph	Unicode Name
U+1171	WI	ㅟ	HANGUL JUNGSEONG WI
U+1172	YU	ㅠ	HANGUL JUNGSEONG YU
U+1173	EU	ㅡ	HANGUL JUNGSEONG EU
U+1174	YI	ㅢ	HANGUL JUNGSEONG YI
U+1175	I	ㅣ	HANGUL JUNGSEONG I
U+11A8	G	ㄱ	HANGUL JONGSEONG KIYEOK
U+11A9	GG	ㄲ	HANGUL JONGSEONG SSANGKIYEOK
U+11AA	GS	ㄳ	HANGUL JONGSEONG KIYEOK-SIOS
U+11AB	N	ㄴ	HANGUL JONGSEONG NIEUN
U+11AC	NJ	ㄵ	HANGUL JONGSEONG NIEUN-CIEUC
U+11AD	NH	ㄶ	HANGUL JONGSEONG NIEUN-HIEUH
U+11AE	D	ㄷ	HANGUL JONGSEONG TIKEUT
U+11AF	L	ㄹ	HANGUL JONGSEONG RIEUL
U+11B0	LG	ㄺ	HANGUL JONGSEONG RIEUL-KIYEOK
U+11B1	LM	ㄻ	HANGUL JONGSEONG RIEUL-MIEUM
U+11B2	LB	ㄼ	HANGUL JONGSEONG RIEUL-PIEUP
U+11B3	LS	ㄽ	HANGUL JONGSEONG RIEUL-SIOS
U+11B4	LT	ㄾ	HANGUL JONGSEONG RIEUL-THIEUTH
U+11B5	LP	ㄿ	HANGUL JONGSEONG RIEUL-PHIEUPH
U+11B6	LH	ㅀ	HANGUL JONGSEONG RIEUL-HIEUH
U+11B7	M	ㅁ	HANGUL JONGSEONG MIEUM
U+11B8	B	ㅂ	HANGUL JONGSEONG PIEUP
U+11B9	BS	ㅄ	HANGUL JONGSEONG PIEUP-SIOS
U+11BA	S	ㅅ	HANGUL JONGSEONG SIOS
U+11BB	SS	ㅆ	HANGUL JONGSEONG SSANGSIOS
U+11BC	NG	ㅇ	HANGUL JONGSEONG IEUNG
U+11BD	J	ㅈ	HANGUL JONGSEONG CIEUC
U+11BE	C	ㅊ	HANGUL JONGSEONG CHIEUCH
U+11BF	K	ㅋ	HANGUL JONGSEONG KHIEUKH
U+11C0	T	ㅌ	HANGUL JONGSEONG THIEUTH
U+11C1	P	ㅍ	HANGUL JONGSEONG PHIEUPH
U+11C2	H	ㅎ	HANGUL JONGSEONG HIEUH

4.5 General Category—Normative in Part

The Unicode Character Database on the CD-ROM defines a General Category for all Unicode characters. This General Category constitutes a partition of the characters into several major classes, such as letters, punctuation, and symbols, and further subclasses for each of the major classes.

Each Unicode character is assigned a General Category value. Each value of the General Category is defined as a two-letter abbreviation, where the first letter gives information about a major class and the second letter designates a subclass of that major class. In each class, the subclass “other” merely collects the remaining characters of the major class. For example, the subclass “No” (Number, other) includes all characters of the Number class that are not a decimal digit or letter. These characters may have little in common besides their membership in the same major class.

Zs, Zl, and Zp are considered format characters, but their membership in the Z (separator) class takes precedence over their membership in the Cf class, because the General Category assigns only a single value to each character.

The General Category is unique, in that some of the values are correlated directly with other normative properties of Unicode characters (numeric status, combining status, or some special properties). Other values are simply informative, provided as an aid to common behavior between Unicode implementations. *Table 4-5* enumerates the values of General Category, with a short description of each value; it is divided between normative and informative values.

Table 4-5. General Category

Normative

Lu = Letter, uppercase
Ll = Letter, lowercase
Lt = Letter, titlecase

Mn = Mark, nonspacing
Mc = Mark, spacing combining
Me = Mark, enclosing

Nd = Number, decimal digit
Nl = Number, letter
No = Number, other

Zs = Separator, space
Zl = Separator, line
Zp = Separator, paragraph

Cc = Other, control
Cf = Other, format
Cs = Other, surrogate
Co = Other, private use
Cn = Other, not assigned

Informative

Lm = Letter, modifier
Lo = Letter, other

Pc = Punctuation, connector
Pd = Punctuation, dash
Ps = Punctuation, open
Pe = Punctuation, close
Pi = Punctuation, initial quote (may behave like Ps or Pe depending on usage)
Pf = Punctuation, final quote (may behave like Ps or Pe depending on usage)
Po = Punctuation, other

Sm = Symbol, math
Sc = Symbol, currency
Sk = Symbol, modifier
So = Symbol, other

A common use of the General Category of a Unicode character is to assist in determination of boundaries in text, as in *Section 5.15, Locating Text Element Boundaries.* Another common use is in determining language identifiers for programming, scripting, and markup, as in *Section 5.16, Identifiers*. This property is also used to support common APIs such as `isLetter()`, `isUppercase()`, and so on.

4.6 Numeric Value—Normative

Numeric value is a normative property of characters that represent *numbers*. This group includes characters such as fractions, subscripts, superscripts, Roman numerals, currency numerators, encircled numbers, and script-specific digits. In many traditional numbering systems, letters are used with a numeric value. Examples include Greek and Hebrew letters as well as Latin letters used in outlines (II.A.1.b). These special cases are not included here as numbers.

Decimal digits form a large subcategory of numbers consisting of those digits that can be used to form decimal-radix numbers. They include script-specific digits, not characters such as Roman numerals (1 + 5 = 15 = fifteen, but I + V = IV = four), subscripts, or superscripts. Numbers other than decimal digits can be used in numerical expressions, but it is up to the user to determine the specialized uses.

The Unicode Standard assigns distinct codes to the forms of digits that are specific to a given script or language. Examples are the digits used with the Arabic script, Chinese numbers, or those of the Indic languages. For naming conventions, see the introduction to *Section 8.2, Arabic*.

Table 4-6 gives the numeric values of Unicode characters that can represent numbers. Some CJK ideographs also have numeric values; those are not included in *Table 4-6*, but are discussed following the table.

Table 4-6. Numeric Properties

Value	Decimal	Number	Name
U+0030	✓	0	DIGIT ZERO
U+0031	✓	1	DIGIT ONE
U+0032	✓	2	DIGIT TWO
U+0033	✓	3	DIGIT THREE
U+0034	✓	4	DIGIT FOUR
U+0035	✓	5	DIGIT FIVE
U+0036	✓	6	DIGIT SIX
U+0037	✓	7	DIGIT SEVEN
U+0038	✓	8	DIGIT EIGHT
U+0039	✓	9	DIGIT NINE
U+00B2		2	SUPERSCRIPT TWO
U+00B3		3	SUPERSCRIPT THREE
U+00B9		1	SUPERSCRIPT ONE
U+00BC		1/4	VULGAR FRACTION ONE QUARTER
U+00BD		1/2	VULGAR FRACTION ONE HALF
U+00BE		3/4	VULGAR FRACTION THREE QUARTERS
U+0660	✓	0	ARABIC-INDIC DIGIT ZERO
U+0661	✓	1	ARABIC-INDIC DIGIT ONE
U+0662	✓	2	ARABIC-INDIC DIGIT TWO
U+0663	✓	3	ARABIC-INDIC DIGIT THREE
U+0664	✓	4	ARABIC-INDIC DIGIT FOUR
U+0665	✓	5	ARABIC-INDIC DIGIT FIVE
U+0666	✓	6	ARABIC-INDIC DIGIT SIX
U+0667	✓	7	ARABIC-INDIC DIGIT SEVEN
U+0668	✓	8	ARABIC-INDIC DIGIT EIGHT
U+0669	✓	9	ARABIC-INDIC DIGIT NINE
U+06F0	✓	0	EXTENDED ARABIC-INDIC DIGIT ZERO
U+06F1	✓	1	EXTENDED ARABIC-INDIC DIGIT ONE
U+06F2	✓	2	EXTENDED ARABIC-INDIC DIGIT TWO
U+06F3	✓	3	EXTENDED ARABIC-INDIC DIGIT THREE

Table 4-6. Numeric Properties (Continued)

Value	Decimal	Number	Name
U+06F4	✓	4	EXTENDED ARABIC-INDIC DIGIT FOUR
U+06F5	✓	5	EXTENDED ARABIC-INDIC DIGIT FIVE
U+06F6	✓	6	EXTENDED ARABIC-INDIC DIGIT SIX
U+06F7	✓	7	EXTENDED ARABIC-INDIC DIGIT SEVEN
U+06F8	✓	8	EXTENDED ARABIC-INDIC DIGIT EIGHT
U+06F9	✓	9	EXTENDED ARABIC-INDIC DIGIT NINE
U+0966	✓	0	DEVANAGARI DIGIT ZERO
U+0967	✓	1	DEVANAGARI DIGIT ONE
U+0968	✓	2	DEVANAGARI DIGIT TWO
U+0969	✓	3	DEVANAGARI DIGIT THREE
U+096A	✓	4	DEVANAGARI DIGIT FOUR
U+096B	✓	5	DEVANAGARI DIGIT FIVE
U+096C	✓	6	DEVANAGARI DIGIT SIX
U+096D	✓	7	DEVANAGARI DIGIT SEVEN
U+096E	✓	8	DEVANAGARI DIGIT EIGHT
U+096F	✓	9	DEVANAGARI DIGIT NINE
U+09E6	✓	0	BENGALI DIGIT ZERO
U+09E7	✓	1	BENGALI DIGIT ONE
U+09E8	✓	2	BENGALI DIGIT TWO
U+09E9	✓	3	BENGALI DIGIT THREE
U+09EA	✓	4	BENGALI DIGIT FOUR
U+09EB	✓	5	BENGALI DIGIT FIVE
U+09EC	✓	6	BENGALI DIGIT SIX
U+09ED	✓	7	BENGALI DIGIT SEVEN
U+09EE	✓	8	BENGALI DIGIT EIGHT
U+09EF	✓	9	BENGALI DIGIT NINE
U+09F4		1	BENGALI CURRENCY NUMERATOR ONE
U+09F5		2	BENGALI CURRENCY NUMERATOR TWO
U+09F6		3	BENGALI CURRENCY NUMERATOR THREE
U+09F7		4	BENGALI CURRENCY NUMERATOR FOUR
U+09F8		-	BENGALI CURRENCY NUMERATOR ONE LESS THAN THE DENOMINATOR
U+09F9		16	BENGALI CURRENCY DENOMINATOR SIXTEEN
U+0A66	✓	0	GURMUKHI DIGIT ZERO
U+0A67	✓	1	GURMUKHI DIGIT ONE
U+0A68	✓	2	GURMUKHI DIGIT TWO
U+0A69	✓	3	GURMUKHI DIGIT THREE
U+0A6A	✓	4	GURMUKHI DIGIT FOUR
U+0A6B	✓	5	GURMUKHI DIGIT FIVE
U+0A6C	✓	6	GURMUKHI DIGIT SIX
U+0A6D	✓	7	GURMUKHI DIGIT SEVEN
U+0A6E	✓	8	GURMUKHI DIGIT EIGHT
U+0A6F	✓	9	GURMUKHI DIGIT NINE
U+0AE6	✓	0	GUJARATI DIGIT ZERO
U+0AE7	✓	1	GUJARATI DIGIT ONE
U+0AE8	✓	2	GUJARATI DIGIT TWO
U+0AE9	✓	3	GUJARATI DIGIT THREE
U+0AEA	✓	4	GUJARATI DIGIT FOUR
U+0AEB	✓	5	GUJARATI DIGIT FIVE
U+0AEC	✓	6	GUJARATI DIGIT SIX
U+0AED	✓	7	GUJARATI DIGIT SEVEN
U+0AEE	✓	8	GUJARATI DIGIT EIGHT
U+0AEF	✓	9	GUJARATI DIGIT NINE
U+0B66	✓	0	ORIYA DIGIT ZERO
U+0B67	✓	1	ORIYA DIGIT ONE
U+0B68	✓	2	ORIYA DIGIT TWO

Table 4-6. Numeric Properties (Continued)

Value	Decimal	Number	Name
U+0B69	✓	3	ORIYA DIGIT THREE
U+0B6A	✓	4	ORIYA DIGIT FOUR
U+0B6B	✓	5	ORIYA DIGIT FIVE
U+0B6C	✓	6	ORIYA DIGIT SIX
U+0B6D	✓	7	ORIYA DIGIT SEVEN
U+0B6E	✓	8	ORIYA DIGIT EIGHT
U+0B6F	✓	9	ORIYA DIGIT NINE
U+0BE7	✓	1	TAMIL DIGIT ONE
U+0BE8	✓	2	TAMIL DIGIT TWO
U+0BE9	✓	3	TAMIL DIGIT THREE
U+0BEA	✓	4	TAMIL DIGIT FOUR
U+0BEB	✓	5	TAMIL DIGIT FIVE
U+0BEC	✓	6	TAMIL DIGIT SIX
U+0BED	✓	7	TAMIL DIGIT SEVEN
U+0BEE	✓	8	TAMIL DIGIT EIGHT
U+0BEF	✓	9	TAMIL DIGIT NINE
U+0BF0		10	TAMIL NUMBER TEN
U+0BF1		100	TAMIL NUMBER ONE HUNDRED
U+0BF2		1,000	TAMIL NUMBER ONE THOUSAND
U+0C66	✓	0	TELUGU DIGIT ZERO
U+0C67	✓	1	TELUGU DIGIT ONE
U+0C68	✓	2	TELUGU DIGIT TWO
U+0C69	✓	3	TELUGU DIGIT THREE
U+0C6A	✓	4	TELUGU DIGIT FOUR
U+0C6B	✓	5	TELUGU DIGIT FIVE
U+0C6C	✓	6	TELUGU DIGIT SIX
U+0C6D	✓	7	TELUGU DIGIT SEVEN
U+0C6E	✓	8	TELUGU DIGIT EIGHT
U+0C6F	✓	9	TELUGU DIGIT NINE
U+0CE6	✓	0	KANNADA DIGIT ZERO
U+0CE7	✓	1	KANNADA DIGIT ONE
U+0CE8	✓	2	KANNADA DIGIT TWO
U+0CE9	✓	3	KANNADA DIGIT THREE
U+0CEA	✓	4	KANNADA DIGIT FOUR
U+0CEB	✓	5	KANNADA DIGIT FIVE
U+0CEC	✓	6	KANNADA DIGIT SIX
U+0CED	✓	7	KANNADA DIGIT SEVEN
U+0CEE	✓	8	KANNADA DIGIT EIGHT
U+0CEF	✓	9	KANNADA DIGIT NINE
U+0D66	✓	0	MALAYALAM DIGIT ZERO
U+0D67	✓	1	MALAYALAM DIGIT ONE
U+0D68	✓	2	MALAYALAM DIGIT TWO
U+0D69	✓	3	MALAYALAM DIGIT THREE
U+0D6A	✓	4	MALAYALAM DIGIT FOUR
U+0D6B	✓	5	MALAYALAM DIGIT FIVE
U+0D6C	✓	6	MALAYALAM DIGIT SIX
U+0D6D	✓	7	MALAYALAM DIGIT SEVEN
U+0D6E	✓	8	MALAYALAM DIGIT EIGHT
U+0D6F	✓	9	MALAYALAM DIGIT NINE
U+0E50	✓	0	THAI DIGIT ZERO
U+0E51	✓	1	THAI DIGIT ONE
U+0E52	✓	2	THAI DIGIT TWO
U+0E53	✓	3	THAI DIGIT THREE
U+0E54	✓	4	THAI DIGIT FOUR
U+0E55	✓	5	THAI DIGIT FIVE
U+0E56	✓	6	THAI DIGIT SIX

Table 4-6. Numeric Properties (Continued)

Value	Decimal	Number	Name
U+0E57	✓	7	THAI DIGIT SEVEN
U+0E58	✓	8	THAI DIGIT EIGHT
U+0E59	✓	9	THAI DIGIT NINE
U+0ED0	✓	0	LAO DIGIT ZERO
U+0ED1	✓	1	LAO DIGIT ONE
U+0ED2	✓	2	LAO DIGIT TWO
U+0ED3	✓	3	LAO DIGIT THREE
U+0ED4	✓	4	LAO DIGIT FOUR
U+0ED5	✓	5	LAO DIGIT FIVE
U+0ED6	✓	6	LAO DIGIT SIX
U+0ED7	✓	7	LAO DIGIT SEVEN
U+0ED8	✓	8	LAO DIGIT EIGHT
U+0ED9	✓	9	LAO DIGIT NINE
U+0F20	✓	0	TIBETAN DIGIT ZERO
U+0F21	✓	1	TIBETAN DIGIT ONE
U+0F22	✓	2	TIBETAN DIGIT TWO
U+0F23	✓	3	TIBETAN DIGIT THREE
U+0F2D	✓	4	TIBETAN DIGIT FOUR
U+0F25	✓	5	TIBETAN DIGIT FIVE
U+0F26	✓	6	TIBETAN DIGIT SIX
U+0F27	✓	7	TIBETAN DIGIT SEVEN
U+0F28	✓	8	TIBETAN DIGIT EIGHT
U+0F29	✓	9	TIBETAN DIGIT NINE
U+0F2A		1/2	TIBETAN DIGIT HALF ONE
U+0F2B		3/2	TIBETAN DIGIT HALF TWO
U+0F2C		5/2	TIBETAN DIGIT HALF THREE
U+0F2D		7/2	TIBETAN DIGIT HALF FOUR
U+0F2E		9/2	TIBETAN DIGIT HALF FIVE
U+0F2F		11/2	TIBETAN DIGIT HALF SIX
U+0F30		13/2	TIBETAN DIGIT HALF SEVEN
U+0F31		15/2	TIBETAN DIGIT HALF EIGHT
U+0F32		17/2	TIBETAN DIGIT HALF NINE
U+0F33		–1/2	TIBETAN DIGIT HALF ZERO
U+1040	✓	0	MYANMAR DIGIT ZERO
U+1041	✓	1	MYANMAR DIGIT ONE
U+1042	✓	2	MYANMAR DIGIT TWO
U+1043	✓	3	MYANMAR DIGIT THREE
U+1044	✓	4	MYANMAR DIGIT FOUR
U+1045	✓	5	MYANMAR DIGIT FIVE
U+1046	✓	6	MYANMAR DIGIT SIX
U+1047	✓	7	MYANMAR DIGIT SEVEN
U+1048	✓	8	MYANMAR DIGIT EIGHT
U+1049	✓	9	MYANMAR DIGIT NINE
U+1369	✓	1	ETHIOPIC DIGIT ONE
U+136A	✓	2	ETHIOPIC DIGIT TWO
U+136B	✓	3	ETHIOPIC DIGIT THREE
U+136C	✓	4	ETHIOPIC DIGIT FOUR
U+136D	✓	5	ETHIOPIC DIGIT FIVE
U+136E	✓	6	ETHIOPIC DIGIT SIX
U+136F	✓	7	ETHIOPIC DIGIT SEVEN
U+1370	✓	8	ETHIOPIC DIGIT EIGHT
U+1371	✓	9	ETHIOPIC DIGIT NINE
U+1372		10	ETHIOPIC NUMBER TEN
U+1373		20	ETHIOPIC NUMBER TWENTY
U+1374		30	ETHIOPIC NUMBER THIRTY
U+1375		40	ETHIOPIC NUMBER FORTY
U+1376		50	ETHIOPIC NUMBER FIFTY

Table 4-6. Numeric Properties (Continued)

Value	Decimal	Number	Name
U+1377		60	ETHIOPIC NUMBER SIXTY
U+1378		70	ETHIOPIC NUMBER SEVENTY
U+1379		80	ETHIOPIC NUMBER EIGHTY
U+137A		90	ETHIOPIC NUMBER NINETY
U+137B		100	ETHIOPIC NUMBER HUNDRED
U+137C		10,000	ETHIOPIC NUMBER TEN THOUSAND
U+16EE		17	RUNIC ARLAUG SYMBOL
U+16EF		18	RUNIC TVIMADUR SYMBOL
U+16F0		19	RUNIC BELGTHOR SYMBOL
U+17E0	✓	0	KHMER DIGIT ZERO
U+17E1	✓	1	KHMER DIGIT ONE
U+17E2	✓	2	KHMER DIGIT TWO
U+17E3	✓	3	KHMER DIGIT THREE
U+17E4	✓	4	KHMER DIGIT FOUR
U+17E5	✓	5	KHMER DIGIT FIVE
U+17E6	✓	6	KHMER DIGIT SIX
U+17E7	✓	7	KHMER DIGIT SEVEN
U+17E8	✓	8	KHMER DIGIT EIGHT
U+17E9	✓	9	KHMER DIGIT NINE
U+1810	✓	0	MONGOLIAN DIGIT ZERO
U+1811	✓	1	MONGOLIAN DIGIT ONE
U+1812	✓	2	MONGOLIAN DIGIT TWO
U+1813	✓	3	MONGOLIAN DIGIT THREE
U+1814	✓	4	MONGOLIAN DIGIT FOUR
U+1815	✓	5	MONGOLIAN DIGIT FIVE
U+1816	✓	6	MONGOLIAN DIGIT SIX
U+1817	✓	7	MONGOLIAN DIGIT SEVEN
U+1818	✓	8	MONGOLIAN DIGIT EIGHT
U+1819	✓	9	MONGOLIAN DIGIT NINE
U+2070		0	SUPERSCRIPT ZERO
U+2074		4	SUPERSCRIPT FOUR
U+2075		5	SUPERSCRIPT FIVE
U+2076		6	SUPERSCRIPT SIX
U+2077		7	SUPERSCRIPT SEVEN
U+2078		8	SUPERSCRIPT EIGHT
U+2079		9	SUPERSCRIPT NINE
U+2080		0	SUBSCRIPT ZERO
U+2081		1	SUBSCRIPT ONE
U+2082		2	SUBSCRIPT TWO
U+2083		3	SUBSCRIPT THREE
U+2084		4	SUBSCRIPT FOUR
U+2085		5	SUBSCRIPT FIVE
U+2086		6	SUBSCRIPT SIX
U+2087		7	SUBSCRIPT SEVEN
U+2088		8	SUBSCRIPT EIGHT
U+2089		9	SUBSCRIPT NINE
U+2153		1/3	VULGAR FRACTION ONE THIRD
U+2154		2/3	VULGAR FRACTION TWO THIRDS
U+2155		1/5	VULGAR FRACTION ONE FIFTH
U+2156		2/5	VULGAR FRACTION TWO FIFTHS
U+2157		3/5	VULGAR FRACTION THREE FIFTHS
U+2158		4/5	VULGAR FRACTION FOUR FIFTHS
U+2159		1/6	VULGAR FRACTION ONE SIXTH
U+215A		5/6	VULGAR FRACTION FIVE SIXTHS
U+215B		1/8	VULGAR FRACTION ONE EIGHTH
U+215C		3/8	VULGAR FRACTION THREE EIGHTHS
U+215D		5/8	VULGAR FRACTION FIVE EIGHTHS
U+215E		7/8	VULGAR FRACTION SEVEN EIGHTHS
U+215F		1	FRACTION NUMERATOR ONE

Table 4-6. Numeric Properties (Continued)

Value	Decimal	Number	Name
U+2160		1	ROMAN NUMERAL ONE
U+2161		2	ROMAN NUMERAL TWO
U+2162		3	ROMAN NUMERAL THREE
U+2163		4	ROMAN NUMERAL FOUR
U+2164		5	ROMAN NUMERAL FIVE
U+2165		6	ROMAN NUMERAL SIX
U+2166		7	ROMAN NUMERAL SEVEN
U+2167		8	ROMAN NUMERAL EIGHT
U+2168		9	ROMAN NUMERAL NINE
U+2169		10	ROMAN NUMERAL TEN
U+216A		11	ROMAN NUMERAL ELEVEN
U+216B		12	ROMAN NUMERAL TWELVE
U+216C		50	ROMAN NUMERAL FIFTY
U+216D		100	ROMAN NUMERAL ONE HUNDRED
U+216E		500	ROMAN NUMERAL FIVE HUNDRED
U+216F		1,000	ROMAN NUMERAL ONE THOUSAND
U+2170		1	SMALL ROMAN NUMERAL ONE
U+2171		2	SMALL ROMAN NUMERAL TWO
U+2172		3	SMALL ROMAN NUMERAL THREE
U+2173		4	SMALL ROMAN NUMERAL FOUR
U+2174		5	SMALL ROMAN NUMERAL FIVE
U+2175		6	SMALL ROMAN NUMERAL SIX
U+2176		7	SMALL ROMAN NUMERAL SEVEN
U+2177		8	SMALL ROMAN NUMERAL EIGHT
U+2178		9	SMALL ROMAN NUMERAL NINE
U+2179		10	SMALL ROMAN NUMERAL TEN
U+217A		11	SMALL ROMAN NUMERAL ELEVEN
U+217B		12	SMALL ROMAN NUMERAL TWELVE
U+217C		50	SMALL ROMAN NUMERAL FIFTY
U+217D		100	SMALL ROMAN NUMERAL ONE HUNDRED
U+217E		500	SMALL ROMAN NUMERAL FIVE HUNDRED
U+217F		1,000	SMALL ROMAN NUMERAL ONE THOUSAND
U+2180		1,000	ROMAN NUMERAL ONE THOUSAND C D
U+2181		5,000	ROMAN NUMERAL FIVE THOUSAND
U+2182		10,000	ROMAN NUMERAL TEN THOUSAND
U+2183		-	ROMAN NUMERAL REVERSED ONE HUNDRED
U+2460		1	CIRCLED DIGIT ONE
U+2461		2	CIRCLED DIGIT TWO
U+2462		3	CIRCLED DIGIT THREE
U+2463		4	CIRCLED DIGIT FOUR
U+2464		5	CIRCLED DIGIT FIVE
U+2465		6	CIRCLED DIGIT SIX
U+2466		7	CIRCLED DIGIT SEVEN
U+2467		8	CIRCLED DIGIT EIGHT
U+2468		9	CIRCLED DIGIT NINE
U+2469		10	CIRCLED NUMBER TEN
U+246A		11	CIRCLED NUMBER ELEVEN
U+246B		12	CIRCLED NUMBER TWELVE
U+246C		13	CIRCLED NUMBER THIRTEEN
U+246D		14	CIRCLED NUMBER FOURTEEN
U+246E		15	CIRCLED NUMBER FIFTEEN
U+246F		16	CIRCLED NUMBER SIXTEEN
U+2470		17	CIRCLED NUMBER SEVENTEEN
U+2471		18	CIRCLED NUMBER EIGHTEEN
U+2472		19	CIRCLED NUMBER NINETEEN
U+2473		20	CIRCLED NUMBER TWENTY
U+2474		1	PARENTHESIZED DIGIT ONE
U+2475		2	PARENTHESIZED DIGIT TWO
U+2476		3	PARENTHESIZED DIGIT THREE
U+2477		4	PARENTHESIZED DIGIT FOUR
U+2478		5	PARENTHESIZED DIGIT FIVE

Table 4-6. Numeric Properties (Continued)

Value	Decimal	Number	Name
U+2479		6	PARENTHESIZED DIGIT SIX
U+247A		7	PARENTHESIZED DIGIT SEVEN
U+247B		8	PARENTHESIZED DIGIT EIGHT
U+247C		9	PARENTHESIZED DIGIT NINE
U+247D		10	PARENTHESIZED NUMBER TEN
U+247E		11	PARENTHESIZED NUMBER ELEVEN
U+247F		12	PARENTHESIZED NUMBER TWELVE
U+2480		13	PARENTHESIZED NUMBER THIRTEEN
U+2481		14	PARENTHESIZED NUMBER FOURTEEN
U+2482		15	PARENTHESIZED NUMBER FIFTEEN
U+2483		16	PARENTHESIZED NUMBER SIXTEEN
U+2484		17	PARENTHESIZED NUMBER SEVENTEEN
U+2485		18	PARENTHESIZED NUMBER EIGHTEEN
U+2486		19	PARENTHESIZED NUMBER NINETEEN
U+2487		20	PARENTHESIZED NUMBER TWENTY
U+2488		1	DIGIT ONE FULL STOP
U+2489		2	DIGIT TWO FULL STOP
U+248A		3	DIGIT THREE FULL STOP
U+248B		4	DIGIT FOUR FULL STOP
U+248C		5	DIGIT FIVE FULL STOP
U+248D		6	DIGIT SIX FULL STOP
U+248E		7	DIGIT SEVEN FULL STOP
U+248F		8	DIGIT EIGHT FULL STOP
U+2490		9	DIGIT NINE FULL STOP
U+2491		10	NUMBER TEN FULL STOP
U+2492		11	NUMBER ELEVEN FULL STOP
U+2493		12	NUMBER TWELVE FULL STOP
U+2494		13	NUMBER THIRTEEN FULL STOP
U+2495		14	NUMBER FOURTEEN FULL STOP
U+2496		15	NUMBER FIFTEEN FULL STOP
U+2497		16	NUMBER SIXTEEN FULL STOP
U+2498		17	NUMBER SEVENTEEN FULL STOP
U+2499		18	NUMBER EIGHTEEN FULL STOP
U+249A		19	NUMBER NINETEEN FULL STOP
U+249B		20	NUMBER TWENTY FULL STOP
U+24EA		0	CIRCLED DIGIT ZERO
U+2776		1	DINGBAT NEGATIVE CIRCLED DIGIT ONE
U+2777		2	DINGBAT NEGATIVE CIRCLED DIGIT TWO
U+2778		3	DINGBAT NEGATIVE CIRCLED DIGIT THREE
U+2779		4	DINGBAT NEGATIVE CIRCLED DIGIT FOUR
U+277A		5	DINGBAT NEGATIVE CIRCLED DIGIT FIVE
U+277B		6	DINGBAT NEGATIVE CIRCLED DIGIT SIX
U+277C		7	DINGBAT NEGATIVE CIRCLED DIGIT SEVEN
U+277D		8	DINGBAT NEGATIVE CIRCLED DIGIT EIGHT
U+277E		9	DINGBAT NEGATIVE CIRCLED DIGIT NINE
U+277F		10	DINGBAT NEGATIVE CIRCLED NUMBER TEN
U+2780		1	DINGBAT CIRCLED SANS-SERIF DIGIT ONE
U+2781		2	DINGBAT CIRCLED SANS-SERIF DIGIT TWO
U+2782		3	DINGBAT CIRCLED SANS-SERIF DIGIT THREE
U+2783		4	DINGBAT CIRCLED SANS-SERIF DIGIT FOUR
U+2784		5	DINGBAT CIRCLED SANS-SERIF DIGIT FIVE
U+2785		6	DINGBAT CIRCLED SANS-SERIF DIGIT SIX
U+2786		7	DINGBAT CIRCLED SANS-SERIF DIGIT SEVEN
U+2787		8	DINGBAT CIRCLED SANS-SERIF DIGIT EIGHT
U+2788		9	DINGBAT CIRCLED SANS-SERIF DIGIT NINE
U+2789		10	DINGBAT CIRCLED SANS-SERIF NUMBER TEN
U+278A		1	DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT ONE
U+278B		2	DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT TWO
U+278C		3	DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT THREE
U+278D		4	DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT FOUR

Table 4-6. Numeric Properties (Continued)

Value	Decimal	Number	Name
U+278E		5	DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT FIVE
U+278F		6	DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT SIX
U+2790		7	DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT SEVEN
U+2791		8	DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT EIGHT
U+2792		9	DINGBAT NEGATIVE CIRCLED SANS-SERIF DIGIT NINE
U+2793		10	DINGBAT NEGATIVE CIRCLED SANS-SERIF NUMBER TEN
U+3007		0	IDEOGRAPHIC NUMBER ZERO
U+3021		1	HANGZHOU NUMERAL ONE
U+3022		2	HANGZHOU NUMERAL TWO
U+3023		3	HANGZHOU NUMERAL THREE
U+3024		4	HANGZHOU NUMERAL FOUR
U+3025		5	HANGZHOU NUMERAL FIVE
U+3026		6	HANGZHOU NUMERAL SIX
U+3027		7	HANGZHOU NUMERAL SEVEN
U+3028		8	HANGZHOU NUMERAL EIGHT
U+3029		9	HANGZHOU NUMERAL NINE
U+3038		10	HANGZHOU NUMERAL TEN
U+3039		20	HANGZHOU NUMERAL TWENTY
U+303A		30	HANGZHOU NUMERAL THIRTY
U+3192		1	IDEOGRAPHIC ANNOTATION ONE MARK
U+3193		2	IDEOGRAPHIC ANNOTATION TWO MARK
U+3194		3	IDEOGRAPHIC ANNOTATION THREE MARK
U+3195		4	IDEOGRAPHIC ANNOTATION FOUR MARK
U+3220		1	PARENTHESIZED IDEOGRAPH ONE
U+3221		2	PARENTHESIZED IDEOGRAPH TWO
U+3222		3	PARENTHESIZED IDEOGRAPH THREE
U+3223		4	PARENTHESIZED IDEOGRAPH FOUR
U+3224		5	PARENTHESIZED IDEOGRAPH FIVE
U+3225		6	PARENTHESIZED IDEOGRAPH SIX
U+3226		7	PARENTHESIZED IDEOGRAPH SEVEN
U+3227		8	PARENTHESIZED IDEOGRAPH EIGHT
U+3228		9	PARENTHESIZED IDEOGRAPH NINE
U+3229		10	PARENTHESIZED IDEOGRAPH TEN
U+3280		1	CIRCLED IDEOGRAPH ONE
U+3281		2	CIRCLED IDEOGRAPH TWO
U+3282		3	CIRCLED IDEOGRAPH THREE
U+3283		4	CIRCLED IDEOGRAPH FOUR
U+3284		5	CIRCLED IDEOGRAPH FIVE
U+3285		6	CIRCLED IDEOGRAPH SIX
U+3286		7	CIRCLED IDEOGRAPH SEVEN
U+3287		8	CIRCLED IDEOGRAPH EIGHT
U+3288		9	CIRCLED IDEOGRAPH NINE
U+3289		10	CIRCLED IDEOGRAPH TEN
U+FF10	✓	0	FULLWIDTH DIGIT ZERO
U+FF11	✓	1	FULLWIDTH DIGIT ONE
U+FF12	✓	2	FULLWIDTH DIGIT TWO
U+FF13	✓	3	FULLWIDTH DIGIT THREE
U+FF14	✓	4	FULLWIDTH DIGIT FOUR
U+FF15	✓	5	FULLWIDTH DIGIT FIVE
U+FF16	✓	6	FULLWIDTH DIGIT SIX
U+FF17	✓	7	FULLWIDTH DIGIT SEVEN
U+FF18	✓	8	FULLWIDTH DIGIT EIGHT
U+FF19	✓	9	FULLWIDTH DIGIT NINE

CJK ideographs from the Unified Repertoire and Ordering also may have numeric values. The primary numeric ideographs are shown in *Table 4-7*. When used to represent numbers

in decimal notation, zero is represented by U+3007. Otherwise, zero is represented by U+96F6.

Table 4-7. Primary Numeric Ideographs

U+96F6	0
U+4E00	1
U+4E8C	2
U+4E09	3
U+56DB	4
U+4E94	5
U+516D	6
U+4E03	7
U+516B	8
U+4E5D	9
U+5341	10
U+767E	100
U+5343	1,000
U+4E07	10,000
U+5104	100,000,000 (10,000 × 10,000)
U+5146	1,000,000,000,000 (10,000 × 10,000 × 10,000)

Ideographic accounting numbers are commonly used on checks and other financial instruments to minimize the possibilities of misinterpretation or fraud in the representation of numerical values. The set of accounting numbers varies somewhat between Japanese, Chinese, and Korean usage. *Table 4-8* gives a fairly complete listing of the known accounting characters. Some of these characters are ideographs with other meanings pressed into service as accounting numbers; others are used only as accounting numbers.

Table 4-8. Ideographs Used as Accounting Numbers

1	U+58F9, U+58F1, U+5F0C[a]
2	U+8CAE[a], U+8D30[a], U+5F10[a], U+5F0D[a]
3	U+53C3, U+53C2, U+53C1[a], U+5F0E[a]
4	U+8086
5	U+4F0D
6	U+9678, U+9646
7	U+67D2[b]
8	U+634C
9	U+7396
10	U+62FE
100	U+4F70[a], U+964C
1,000	U+4EDF
10,000	U+842C

a. These characters are used *only* as accounting numbers, and have no other meaning.
b. In Japan, U+67D2 is also pronounced *urusi*, meaning "lacquer," and is treated as a variant of the standard character for "lacquer" U+6F06.

4.7 Mirrored—Normative

Mirrored is a normative property of characters such as parentheses, whose images are mirrored horizontally in text that is laid out from right to left. For example, U+0028 LEFT PARENTHESIS is interpreted as *opening parenthesis*; in a left-to-right context it will appear as "(", while in a right-to-left context it will appear as the mirrored glyph ")". The list of mirrored characters appears in *Table 4-9*. Note that mirroring is not limited to paired characters, but that any character with the mirrored property will need two mirrored glyphs. This requirement is necessary to render the character properly in a bidirectional context.

Table 4-9. Mirrored Characters

Value	Name
U+0028	LEFT PARENTHESIS
U+0029	RIGHT PARENTHESIS
U+003C	LESS-THAN SIGN
U+003E	GREATER-THAN SIGN
U+005B	LEFT SQUARE BRACKET
U+005D	RIGHT SQUARE BRACKET
U+007B	LEFT CURLY BRACKET
U+007D	RIGHT CURLY BRACKET
U+2045	LEFT SQUARE BRACKET WITH QUILL
U+2046	RIGHT SQUARE BRACKET WITH QUILL
U+207D	SUPERSCRIPT LEFT PARENTHESIS
U+207E	SUPERSCRIPT RIGHT PARENTHESIS
U+208D	SUBSCRIPT LEFT PARENTHESIS
U+208E	SUBSCRIPT RIGHT PARENTHESIS
U+2201	COMPLEMENT
U+2202	PARTIAL DIFFERENTIAL
U+2203	THERE EXISTS
U+2204	THERE DOES NOT EXIST
U+2208	ELEMENT OF
U+2209	NOT AN ELEMENT OF
U+220A	SMALL ELEMENT OF
U+220B	CONTAINS AS MEMBER
U+220C	DOES NOT CONTAIN AS MEMBER
U+220D	SMALL CONTAINS AS MEMBER
U+2211	N-ARY SUMMATION
U+2215	DIVISION SLASH
U+2216	SET MINUS
U+221A	SQUARE ROOT
U+221B	CUBE ROOT
U+221C	FOURTH ROOT
U+221D	PROPORTIONAL TO
U+221F	RIGHT ANGLE
U+2220	ANGLE
U+2221	MEASURED ANGLE
U+2222	SPHERICAL ANGLE
U+2224	DOES NOT DIVIDE
U+2226	NOT PARALLEL TO
U+222B	INTEGRAL
U+222C	DOUBLE INTEGRAL
U+222D	TRIPLE INTEGRAL
U+222E	CONTOUR INTEGRAL
U+222F	SURFACE INTEGRAL
U+2230	VOLUME INTEGRAL
U+2231	CLOCKWISE INTEGRAL
U+2232	CLOCKWISE CONTOUR INTEGRAL
U+2233	ANTICLOCKWISE CONTOUR INTEGRAL
U+2239	EXCESS
U+223B	HOMOTHETIC
U+223C	TILDE OPERATOR
U+223D	REVERSED TILDE
U+223E	INVERTED LAZY S
U+223F	SINE WAVE
U+2240	WREATH PRODUCT
U+2241	NOT TILDE
U+2242	MINUS TILDE
U+2243	ASYMPTOTICALLY EQUAL TO
U+2244	NOT ASYMPTOTICALLY EQUAL TO
U+2245	APPROXIMATELY EQUAL TO
U+2246	APPROXIMATELY BUT NOT ACTUALLY EQUAL TO
U+2247	NEITHER APPROXIMATELY NOR ACTUALLY EQUAL TO
U+2248	ALMOST EQUAL TO

Table 4-9. Mirrored Characters (Continued)

Value	Name
U+2249	NOT ALMOST EQUAL TO
U+224A	ALMOST EQUAL OR EQUAL TO
U+224B	TRIPLE TILDE
U+224C	ALL EQUAL TO
U+2252	APPROXIMATELY EQUAL TO OR THE IMAGE OF
U+2253	IMAGE OF OR APPROXIMATELY EQUAL TO
U+2254	COLON EQUALS
U+2255	EQUALS COLON
U+225F	QUESTIONED EQUAL TO
U+2260	NOT EQUAL TO
U+2262	NOT IDENTICAL TO
U+2264	LESS-THAN OR EQUAL TO
U+2265	GREATER-THAN OR EQUAL TO
U+2266	LESS-THAN OVER EQUAL TO
U+2267	GREATER-THAN OVER EQUAL TO
U+2268	LESS-THAN BUT NOT EQUAL TO
U+2269	GREATER-THAN BUT NOT EQUAL TO
U+226A	MUCH LESS-THAN
U+226B	MUCH GREATER-THAN
U+226E	NOT LESS-THAN
U+226F	NOT GREATER-THAN
U+2270	NEITHER LESS-THAN NOR EQUAL TO
U+2271	NEITHER GREATER-THAN NOR EQUAL TO
U+2272	LESS-THAN OR EQUIVALENT TO
U+2273	GREATER-THAN OR EQUIVALENT TO
U+2274	NEITHER LESS-THAN NOR EQUIVALENT TO
U+2275	NEITHER GREATER-THAN NOR EQUIVALENT TO
U+2276	LESS-THAN OR GREATER-THAN
U+2277	GREATER-THAN OR LESS-THAN
U+2278	NEITHER LESS-THAN NOR GREATER-THAN
U+2279	NEITHER GREATER-THAN NOR LESS-THAN
U+227A	PRECEDES
U+227B	SUCCEEDS
U+227C	PRECEDES OR EQUAL TO
U+227D	SUCCEEDS OR EQUAL TO
U+227E	PRECEDES OR EQUIVALENT TO
U+227F	SUCCEEDS OR EQUIVALENT TO
U+2280	DOES NOT PRECEDE
U+2281	DOES NOT SUCCEED
U+2282	SUBSET OF
U+2283	SUPERSET OF
U+2284	NOT A SUBSET OF
U+2285	NOT A SUPERSET OF
U+2286	SUBSET OF OR EQUAL TO
U+2287	SUPERSET OF OR EQUAL TO
U+2288	NEITHER A SUBSET OF NOR EQUAL TO
U+2289	NEITHER A SUPERSET OF NOR EQUAL TO
U+228A	SUBSET OF WITH NOT EQUAL TO
U+228B	SUPERSET OF WITH NOT EQUAL TO
U+228C	MULTISET
U+228F	SQUARE IMAGE OF
U+2290	SQUARE ORIGINAL OF
U+2291	SQUARE IMAGE OF OR EQUAL TO
U+2292	SQUARE ORIGINAL OF OR EQUAL TO
U+2298	CIRCLED DIVISION SLASH
U+22A2	RIGHT TACK
U+22A3	LEFT TACK
U+22A6	ASSERTION
U+22A7	MODELS
U+22A8	TRUE
U+22A9	FORCES

Table 4-9. Mirrored Characters (Continued)

Value	Name
U+22AA	TRIPLE VERTICAL BAR RIGHT TURNSTILE
U+22AB	DOUBLE VERTICAL BAR DOUBLE RIGHT TURNSTILE
U+22AC	DOES NOT PROVE
U+22AD	NOT TRUE
U+22AE	DOES NOT FORCE
U+22AF	NEGATED DOUBLE VERTICAL BAR DOUBLE RIGHT TURNSTILE
U+22B0	PRECEDES UNDER RELATION
U+22B1	SUCCEEDS UNDER RELATION
U+22B2	NORMAL SUBGROUP OF
U+22B3	CONTAINS AS NORMAL SUBGROUP
U+22B4	NORMAL SUBGROUP OF OR EQUAL TO
U+22B5	CONTAINS AS NORMAL SUBGROUP OR EQUAL TO
U+22B6	ORIGINAL OF
U+22B7	IMAGE OF
U+22B8	MULTIMAP
U+22BE	RIGHT ANGLE WITH ARC
U+22BF	RIGHT TRIANGLE
U+22C9	LEFT NORMAL FACTOR SEMIDIRECT PRODUCT
U+22CA	RIGHT NORMAL FACTOR SEMIDIRECT PRODUCT
U+22CB	LEFT SEMIDIRECT PRODUCT
U+22CC	RIGHT SEMIDIRECT PRODUCT
U+22CD	REVERSED TILDE EQUALS
U+22D0	DOUBLE SUBSET
U+22D1	DOUBLE SUPERSET
U+22D6	LESS-THAN WITH DOT
U+22D7	GREATER-THAN WITH DOT
U+22D8	VERY MUCH LESS-THAN
U+22D9	VERY MUCH GREATER-THAN
U+22DA	LESS-THAN EQUAL TO OR GREATER-THAN
U+22DB	GREATER-THAN EQUAL TO OR LESS-THAN
U+22DC	EQUAL TO OR LESS-THAN
U+22DD	EQUAL TO OR GREATER-THAN
U+22DE	EQUAL TO OR PRECEDES
U+22DF	EQUAL TO OR SUCCEEDS
U+22E0	DOES NOT PRECEDE OR EQUAL
U+22E1	DOES NOT SUCCEED OR EQUAL
U+22E2	NOT SQUARE IMAGE OF OR EQUAL TO
U+22E3	NOT SQUARE ORIGINAL OF OR EQUAL TO
U+22E4	SQUARE IMAGE OF OR NOT EQUAL TO
U+22E5	SQUARE ORIGINAL OF OR NOT EQUAL TO
U+22E6	LESS-THAN BUT NOT EQUIVALENT TO
U+22E7	GREATER-THAN BUT NOT EQUIVALENT TO
U+22E8	PRECEDES BUT NOT EQUIVALENT TO
U+22E9	SUCCEEDS BUT NOT EQUIVALENT TO
U+22EA	NOT NORMAL SUBGROUP OF
U+22EB	DOES NOT CONTAIN AS NORMAL SUBGROUP
U+22EC	NOT NORMAL SUBGROUP OF OR EQUAL TO
U+22ED	DOES NOT CONTAIN AS NORMAL SUBGROUP OR EQUAL
U+22F0	UP RIGHT DIAGONAL ELLIPSIS
U+22F1	DOWN RIGHT DIAGONAL ELLIPSIS
U+2308	LEFT CEILING
U+2309	RIGHT CEILING
U+230A	LEFT FLOOR
U+230B	RIGHT FLOOR
U+2320	TOP HALF INTEGRAL
U+2321	BOTTOM HALF INTEGRAL
U+2329	LEFT-POINTING ANGLE BRACKET
U+232A	RIGHT-POINTING ANGLE BRACKET
U+3008	LEFT ANGLE BRACKET
U+3009	RIGHT ANGLE BRACKET
U+300A	LEFT DOUBLE ANGLE BRACKET

Table 4-9. Mirrored Characters (Continued)

Value	Name
U+300B	RIGHT DOUBLE ANGLE BRACKET
U+300C	LEFT CORNER BRACKET
U+300D	RIGHT CORNER BRACKET
U+300E	LEFT WHITE CORNER BRACKET
U+300F	RIGHT WHITE CORNER BRACKET
U+3010	LEFT BLACK LENTICULAR BRACKET
U+3011	RIGHT BLACK LENTICULAR BRACKET
U+3014	LEFT TORTOISE SHELL BRACKET
U+3015	RIGHT TORTOISE SHELL BRACKET
U+3016	LEFT WHITE LENTICULAR BRACKET
U+3017	RIGHT WHITE LENTICULAR BRACKET
U+3018	LEFT WHITE TORTOISE SHELL BRACKET
U+3019	RIGHT WHITE TORTOISE SHELL BRACKET
U+301A	LEFT WHITE SQUARE BRACKET
U+301B	RIGHT WHITE SQUARE BRACKET

4.8 Unicode 1.0 Names

The *Unicode 1.0 character name* is an informative property of the characters defined in Version 1.0 of the Unicode Standard. The names of Unicode characters were changed in the process of merging the standard with ISO/IEC 10646. The Version 1.0 character names can be obtained from the CD-ROM accompanying the standard or from the ftp site. See also *Appendix D, Changes from Unicode Version 2.0*. Where the Version 1.0 character name provides additional useful information, it is listed in *Chapter 14, Code Charts*. For example, U+00B6 PILCROW SIGN has its Version 1.0 name, PARAGRAPH SIGN, listed for clarity.

4.9 Mathematical Property

The mathematical property is an informative property of characters that are used as operators in mathematical formulas. The mathematical property may be useful in algorithms that deal with the display of mathematical text and formulas. However, a number of these characters have multiple usages and may occur with nonmathematical semantics. For example, U+002D HYPHEN-MINUS may also be used as a hyphen—and not as a mathematical minus sign. Other characters, including some alphabetic, numeric, punctuation, spaces, arrows, and geometric shapes, are used in mathematical expressions as well, but are even more dependent on the context for their identification. The Unicode characters in the following lists have the mathematical property.

Characters with the math property and the Sm General Category:

```
002B, 003C..003E, 007C, 007E, 00AC, 00B1, 00D7, 00F7, 2044,
207A..207C, 208A..208C, 2190..2194, 219A..219B, 21A0, 21A3, 21A6,
21AE, 21CE..21CF, 21D2, 21D4, 2200..22F1, 2308..230B, 2320..2321,
25B7, 25C1, 266F, FB29, FE62, FE64..FE66, FF0B, FF1C..FF1E, FF5C,
FF5E, FFE2, FFE9..FFEC
```

Characters with the math property and other General Category values:

```
0028..002A, 002D, 002F, 005B..005E, 007B, 007D, 2016, 2032..2034,
207D..207E, 208D..208E, 20D0..20DC, 20E1, 2329..232A, 300A..300B,
301A..301B, FE35..FE38, FE59..FE5C, FE61, FE63, FE68, FF08..FF0A,
FF0D, FF0F, FF3B..FF3E, FF5B, FF5D
```

4.10 Letters and Other Useful Properties

The CD-ROM that accompanies the Unicode Standard contains data files that list other useful, informative properties of Unicode characters. The full list of those properties can be found in the data files; see, in particular, PropList.txt. This section highlights some of those properties that have a bearing on such implementation issues as parsing of identifiers. (See also *Section 5.16, Identifiers.*)

Computer language standards often characterize identifiers as consisting of letters, syllables, ideographs, and digits, but do not specify exactly what a "letter," "syllable," "ideograph," or "digit" is, leaving the definitions implicitly either to a character encoding standard or to a locale specification. The large scope of the Unicode Standard means that it includes many writing systems for which these distinctions are not as self-evident as they may once have been for systems designed to work primarily for Western European languages and Japanese. In particular, while the Unicode Standard includes various "alphabets" and "syllabaries," it also includes writing systems that fall somewhere in between. As a result, no attempt is made to draw a sharp property distinction between letters and syllables.

Letter. This informative property applies to characters that are used to write words. This group includes characters such as capital letters, small letters, ideographs, hangul, and spacing modifier letters. Combining marks generally assume the letter property of the preceding base character. For example, when searching for word boundaries, combining characters don't break from previous letters. The letter property mappings can be obtained from the CD-ROM accompanying the standard.

Alphabetic. The alphabetic property is an informative property of the primary units of alphabets and/or syllabaries, whether combining or noncombining. Included in this group would be composite characters that are canonical equivalents to a combining character sequence of an alphabetic base character plus one or more combining characters; letter digraphs; contextual variants of alphabetic characters; ligatures of alphabetic characters; contextual variants of ligatures; modifier letters; letterlike symbols that are compatibility equivalents of single alphabetic letters; and miscellaneous letter elements. Notably, U+00AA FEMININE ORDINAL INDICATOR and U+00BA MASCULINE ORDINAL INDICATOR are simply abbreviatory forms involving a Latin letter and should be considered alphabetic rather than nonalphabetic symbols.

Ideographic. The ideographic property is an informative property of the Unified CJK Ideograph set (U+4E00..U+9FA5); the CJK Ideograph Extension A set (U+3400..U+4DB5); the CJK Compatibility Ideograph set (U+F900..U+FA2D); U+3007 IDEOGRAPHIC NUMBER ZERO; U+3006 IDEOGRAPHIC CLOSING MARK; and the Hangzhou-style numerals (U+3021.. U+3029, U+3038..U+303A).

This PDF file is an excerpt from *The Unicode Standard, Version 3.0*, issued by the Unicode Consortium and published by Addison-Wesley. The material has been modified slightly for this online edition, however the PDF files have not been modified to reflect the corrections found on the Updates and Errata page (see http://www.unicode.org/unicode/uni2errata/UnicodeErrata.html). More recent versions of the Unicode standard exist (see http://www.unicode.org/unicode/standard/versions/).

Dai Kan-Wa Jiten used as the source of reference Kanji codes was written by Tetsuji Morohashi and published by Taishukan Shoten.

ISBN 0-201-61633-5

This book is set in Minion, designed by Rob Slimbach at Adobe Systems, Inc. It was typeset using FrameMaker 5.5 running under Windows NT. ASMUS, Inc. created custom software for chart layout. The Han radical-stroke index was typeset by Apple Computer, Inc. The following companies and organizations supplied fonts:

Apple Computer, Inc.
Atelier Fluxus Virus
Beijing Zhong Yi (Zheng Code) Electronics Company
DecoType, Inc.
IBM Corporation
Monotype Typography, Inc.
Microsoft Corporation
Peking University Founder Group Corporation
Production First Software

Additional fonts were supplied by individuals as listed in the *Acknowledgments*.

The publisher offers discounts on this book when ordered in quantity for special sales. For more information please contact:

Corporate, Government, and Special Sales
Addison Wesley Longman, Inc.
One Jacob Way
Reading, Massachusetts 01867

Visit A-W on the Web: http://www.awl.com/cseng/

First printing, January 2000.