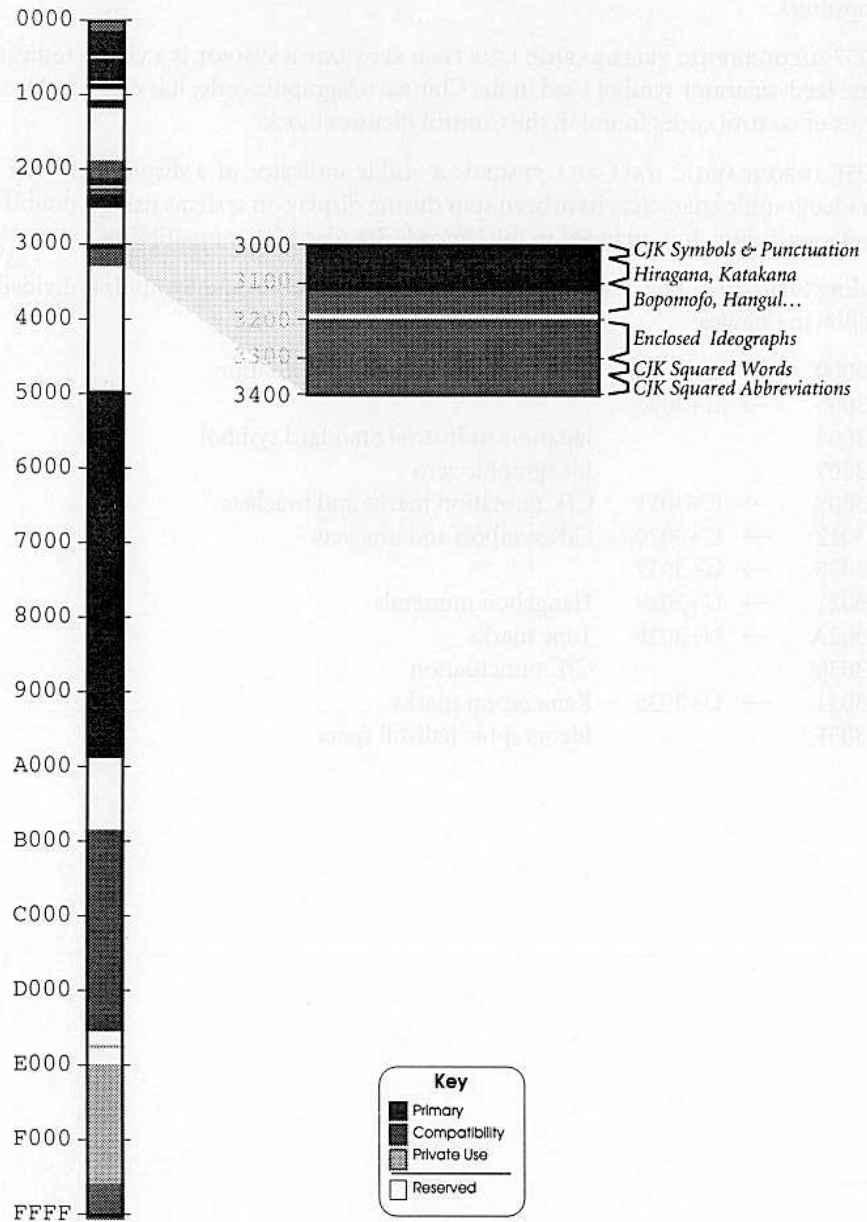


6.3 CJK Phonetics and Symbols Area

The CJK Phonetics and Symbols Area of the Unicode Standard includes the encoding of phonetic characters, punctuation marks and symbols used in the CJK (Chinese, Japanese, Korean) writing systems (see Figure 6-21).

Figure 6-21. CJK Misc



CJK Symbols and Punctuation: U+3000—U+303F

This block encodes punctuation marks and symbols used primarily by writing systems that employ Han ideographs. Most of these characters are found in East Asian standards.

U+3000 IDEOGRAPHIC SPACE is provided for compatibility. It is a fixed-width space appropriate for use with an ideographic font. U+301C WAVE DASH and U+3030 WAVY DASH are special forms of dashes found in East Asian character standards. (For a list of other space and dash characters in the Unicode Standard, see the General Punctuation character block description.)

U+3037 IDEOGRAPHIC TELEGRAPHIC LINE FEED SEPARATOR SYMBOL is a visible indicator of the line feed separator symbol used in the Chinese telegraphic code; it is comparable to the pictures of control codes found in the Control Pictures block.

U+303F IDEOGRAPHIC HALF-FILL SPACE is a visible indicator of a display cell filler used when ideographic characters have been split during display on systems using a double-byte character encoding. It is included in the Unicode Standard for compatibility.

Encoding Structure. The character block for CJK Punctuation and Symbols is divided into the following ranges:

| | | |
|--------|-----------|-------------------------------------|
| U+3000 | → U+3003, | Ideographic space and punctuation |
| U+3005 | → U+3006 | |
| U+3004 | | Japanese Industrial Standard symbol |
| U+3007 | | Ideographic zero |
| U+3008 | → U+3011 | CJK quotation marks and brackets |
| U+3012 | → U+3020, | CJK symbols and brackets |
| U+3036 | → U+3037 | |
| U+3021 | → U+3029 | Hangzhou numerals |
| U+302A | → U+302F | Tone marks |
| U+3030 | | CJK punctuation |
| U+3031 | → U+3035 | Kana repeat marks |
| U+303F | | Ideographic half-fill space |

Hiragana: U+3040—U+309F

Hiragana is the cursive syllabary used to write Japanese words phonetically and also to write sentence particles and inflectional endings. Hiragana is commonly used as well to indicate the pronunciation of Japanese words. Hiragana syllables are phonetically equivalent to corresponding Katakana syllables.

Standards. The Hiragana block is based on the JIS X 0208-1990 standard, extended by the non-standard syllable U+3094 *vu*, which is included in some Japanese corporate standards.

Combining Marks. Hiragana and the related script Katakana use U+3099 *COMBINING KATAKANA-HIRAGANA VOICED SOUND MARK* and U+309A *COMBINING KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK* to generate voiced and semi-voiced syllables from the base syllables, respectively. All common precomposed combinations of base syllable forms using these marks are already encoded as characters, and use of these precomposed forms is the predominant JIS usage. These combining marks must follow the base character to which they apply. As most implementations and JIS standard treat these as spacing characters, the Unicode Standard also contains two corresponding non-combining (spacing) marks at U+309B and U+309C.

Iteration Marks. The two characters U+309D *HIRAGANA ITERATION MARK* and U+309E *HIRAGANA VOICED ITERATION MARK* are punctuation-like characters that denote the iteration (repetition) of a previous syllable according to whether the repeated syllable has an unvoiced or voiced consonant, respectively.

Encoding Structure. The character block for the Hiragana script is divided into the following ranges:

| | | | |
|--------|---|--------|---|
| U+3041 | → | U+3093 | Hiragana syllabary |
| U+3094 | | | Extended Hiragana syllable form |
| U+3099 | → | U+309A | Combining (non-spacing) diacritical marks |
| U+309B | → | U+309C | Non-Combining (spacing) diacritical marks |
| U+309D | → | U+309E | Iteration marks |

Katakana: U+30A0—U+30FF

Katakana is the non-cursive syllabary used to write non-Japanese (usually Western) words phonetically in Japanese. Katakana is also used as well to write Japanese words with visual emphasis. Katakana syllables are phonetically equivalent to corresponding Hiragana syllables. Katakana contains two characters, U+30F5 KATAKANA LETTER SMALL KA and U+30F6 KATAKANA LETTER SMALL KE, which have no direct correspondent in Hiragana; these are used in special Japanese spelling conventions (for example, the spelling of placenames which include archaic Japanese connective particles).

Standards. The Katakana block is based on the JIS X 0208-1990 standard.

Punctuation-like Characters. U+30FB KATAKANA MIDDLE DOT is used to separate words when writing non-Japanese phrases. U+30FC KATAKANA-HIRAGANA PROLONGED SOUND MARK is used predominantly with Katakana and occasionally with Hiragana in order to denote a lengthened vowel of the previously written syllable. The two iteration marks, U+30FD KATAKANA ITERATION MARK and U+30FE KATAKANA VOICED ITERATION MARK serve the same function in Katakana writing that the two Hiragana iteration marks serve in Hiragana writing.

Encoding Structure. The character block for the Katakana script is divided into the following ranges:

| | | | |
|--------|---|--------|--|
| U+30A1 | → | U+30F6 | Katakana syllabary |
| U+30F7 | → | U+30FA | Extended Katakana syllable forms |
| U+30FB | → | U+30FC | Punctuation and vowel lengthening mark |
| U+30FD | → | U+30FE | Iteration marks |

Bopomofo: U+3100—U+312F

Bopomofo are a set of characters used to annotate or teach the phonetics of Chinese, primarily the standard Mandarin language. They are used in dictionaries and teaching materials, not in the actual writing of Chinese text. The formal Chinese names for this alphabet are *Zhuyin-Zimu* (“phonetic alphabet”) or *Zhuyin-Fuhao* (“phonetic symbols”), but the informal term Bopomofo (analogous to “ABCs”) provides a more serviceable English name and is also used in China. The Bopomofo were developed as part of a populist literacy campaign following the 1911 revolution; thus they are acceptable to all branches of modern Chinese culture, although in the People’s Republic of China their function has been largely taken over by the Pinyin romanization system.

Standards. The standard Mandarin set of Bopomofo are included in the People’s Republic of China standard GB 2312 and in the Republic of China (Taiwan) standard CNS 11643.

Mandarin Tone Marks. Small modifier letters used to indicate the five Mandarin tones are part of the Bopomofo system, but in the Unicode Standard they have been unified into the Modifier Letter range, as follows:

| | | |
|-------------|--------|------------------------------|
| first tone | U+02C9 | MODIFIER LETTER MACRON |
| second tone | U+02CA | MODIFIER LETTER ACUTE ACCENT |
| third tone | U+02C7 | CARON |
| fourth tone | U+02CB | MODIFIER LETTER GRAVE ACCENT |
| light tone | U+02D9 | DOT ABOVE |

Standard Mandarin Bopomofo. The order of Bopomofo letters is standard worldwide. The code offset of the first letter U+3105 BOPOMOFO LETTER B from a multiple of 16 is included to match the offset in the ISO-registered standard GB 2312. The character U+3127 BOPOMOFO LETTER I is usually written as a horizontal stroke when the Bopomofo text is set vertically; in the Unicode Standard this is considered to be a rendering variation; the variant is not assigned a separate character code.

Non-Mandarin Letters. These are very rarely used but are included for completeness. There are no standard Bopomofo letters for the phonetics of Cantonese or other dialects.

Encoding Structure. The character block for Bopomofo is divided into the following ranges:

| | | | |
|--------|---|--------|--------------------------------|
| U+3105 | → | U+3129 | Standard Mandarin Bopomofo |
| U+312A | → | U+312C | Dialect (non-Mandarin) letters |

Hangul Compatibility Jamo: U+3130—U+318F

This block consists of spacing, non-conjoining Hangul consonant and vowel (jamo) elements. These characters are provided solely for compatibility with the KS C 5601 standard. Unlike the characters found in the Hangul Jamo block (U+1100 → U+11FF), the jamo characters in this block have no conjoining semantics.

The characters of this block are considered to be full-width forms in contrast with the Half-width Hangul Compatibility Jamo found at U+FFA0 → U+FFDF.

Standards. The Unicode Standard follows KS C 5601 for Hangul Jamo elements.

Encoding Structure. The character block for Hangul Elements is divided into the following ranges:

| | | |
|--------|----------|-----------------------|
| U+3131 | → U+3163 | Modern Jamo elements |
| U+3164 | | Hangul fill character |
| U+3165 | → U+318E | Archaic Jamo elements |

Kanbun: U+3190—U+319F

This block contains a set of Kanbun marks used in Japanese texts to indicate the Japanese reading order of classical Chinese texts. They are not encoded in any current character encoding standards but are widely used in literature. They are typically written in an aino-tation style to the left of each line of vertically-rendered Chinese text.

Encoding Structure. The character block for Kanbun is divided into the following ranges:

U+3190 → U+319F Kanbun marks

Enclosed CJK Letters and Months: U+3200—U+32FF

Standards. This block provides mapping for all the enclosed Hangul elements from Korean standard KS C 5601 as well as parenthesized ideographic characters from JIS X 0208-1990 standard, CNS 11643, and several corporate registries.

Encoding Structure. The character block for CJK parenthesized letters and ideographs is divided into the following ranges:

| | | | |
|--------|---|--------|--|
| U+3200 | → | U+320D | Parenthesized Hangul Jamo elements |
| U+320E | → | U+321C | Parenthesized Hangul syllables |
| U+3220 | → | U+3229 | Parenthesized number ideographs |
| U+322A | → | U+3243 | Parenthesized ideographs |
| U+3260 | → | U+326D | Circled Hangul Jamo elements |
| U+326E | → | U+327B | Circled Hangul syllables |
| U+327F | | | Korean standard symbol |
| U+3280 | → | U+3289 | Circled number ideographs |
| U+328A | → | U+32B0 | Circled ideographs |
| U+32C0 | → | U+32CB | Ideographic telegraph symbols for months |
| U+32D0 | → | U+32FE | Circled Katakana |

CJK Compatibility: U+3300—U+33FF

CJK squared Katakana words are Katakana-spelled words that fill a single display cell (em-square) when intermixed with CJK ideographs. Likewise, squared Latin abbreviation symbols are designed to fill a single character position when mixed with CJK ideographs.

These characters are provided solely for compatibility with existing character encoding standards.

Standards. CJK Compatibility characters are derived from the KS C 5601 and CNS 11643 national standards, and from various company registries.

Japanese Era Names. The Japanese era names refer to the following dates:

| | | |
|--------|-------------------------|-------------------------|
| U+337B | SQUARE ERA NAME HEISEI | 1989/1/7 to present day |
| U+337C | SQUARE ERA NAME SYOUWA | 1926/12/24 to 1989/1/6 |
| U+337D | SQUARE ERA NAME TAISYOU | 1912/7/29 to 1926/12/23 |
| U+337E | SQUARE ERA NAME MEIZI | 1867 to 1912/7/28 |

Encoding Structure. The character block for CJK Compatibility characters is divided into the following ranges:

| | | |
|--------|----------|--|
| U+3300 | → U+3357 | Squared symbolic Katakana words |
| U+3358 | → U+3370 | Ideographic telegraph symbols for hours |
| U+3371 | → U+3376 | Squared Latin abbreviation symbols |
| U+337B | → U+337E | Japanese era names |
| U+337F | | Squared Ideographic word ('corporation') |
| U+3380 | → U+33DD | Squared Latin abbreviation symbols |
| U+33E0 | → U+33FE | Ideographic telegraph symbols for days |