

## Chapter 6

# *Character Block Descriptions*

The character codes in the character charts constitute the core of the Unicode Standard. The characters cover modern communication and computer usage in the major scripts of the world. The coverage consists of over 2,700 alphabetic (or syllabic) letters, over 1,300 textual symbols, over 800 CJK phonetics and symbols, 11,172 Korean Hangul syllables, and over 21,000 Han (East Asian ideographic) characters.

Characters are organized into related groups called *blocks*. The following block descriptions are written so that they can be read alone, but by themselves they are not sufficient to implement the Unicode character encoding. It is also vital to understand the general architecture of the Unicode Standard (see *Chapter 2, General Structure*, and *Chapter 3, Conformance*).

The character blocks and the block descriptions are ordered by their code range location. Thus, the character blocks are ordered in the following sequence: the General Scripts Area, the Symbols Area, the CJK Phonetics and Symbols Area, the CJK Ideographs Area, the Hangul Syllables Area, the Surrogates Area, the Private Use Area, and the Compatibility Area and Specials. (See Figure 2-5 for an overview of scripts included in this version of the Unicode Standard.)

In the Unicode Standard, character blocks generally contain characters from a single script. In many cases a script is fully represented in its character block. There are, however, important exceptions, most notably in the area of punctuation characters. Punctuation characters occur in several widely separated places in the character blocks, including Basic Latin, Latin-1 Supplement, General Punctuation, and CJK Symbols and Punctuation, as well as occasional characters in character blocks for specific scripts. It is intended that punctuation characters, for example U+002C COMMA or U+2022 BULLET, be encoded only once, rather than being encoded again and again for particular scripts; such general purpose punctuation may be used for any script or mixture of scripts.