
Vendor Character Set Standards

This appendix is not included in the printed version of this book, and is instead available as a downloadable and printable PDF file. As new material becomes available, the PDF file will be updated accordingly.*

The material currently included is directly excerpted from the first edition, and an updated appendix for the second edition will be made available as that material becomes available.

* <http://examples.oreilly.com/9780596514471/cjkvip2e-appE.pdf>

C

Vendor Character Set Standards

The material presented in this appendix supplements Chapter 3, *Character Set Standards*, by including detailed information on vendor CJKV character set standards. This appendix is primarily intended as reference material in the event that you need to deal with one of the included character set standards. The character set tables that are included supplement those that appear later in this book. This book's *Bibliography* has information on the documentation for some of the character set standards covered in this appendix.

Nearly all of the vendor character sets described in this appendix are based on one or more national standards, and usually provide additional symbols and Chinese characters. In addition, you will learn that many vendor-defined Chinese characters can be found in supplemental national character set standards, such as in Japan's JIS X 0212-1990 for Japanese vendor character sets.

The vendor character set standards covered in this appendix do not represent an exhaustive list—nearly every major computer manufacturer that does business in one or more CJKV locales has developed its own character set standard. This material shows you not only how diverse character sets can be, but also how they are not fully compatible with one another nor with the national character set standards covered in Chapter 3.

The Ideographic Rapporteur Group* (IRG; formerly CJK-JRG) has diligently worked on adding 6,582 more Chinese characters to Unicode. These additional characters have been approved and have mappings. Some of the characters you find in this appendix may soon become part of the Unicode character set (if they are not included already).

* <http://www.cs.cubk.edu.hk/~irg/>

Many pages of this chapter are dedicated to IBM's Chinese, Japanese, and Korean character sets. For more detailed information on IBM's character sets, to include machine-readable mapping tables on CD-ROM, I strongly suggest that you buy IBM's *Character Data Representation Architecture Reference and Registry* (1995, IBM part number SC09-2190-00). Page 982 in Appendix T, *Software and Document Sources*, provides phone numbers and a URL for ordering this IBM publication.

Chinese Vendor Character Sets—China

Besides the GBK character set, which is an extended version of GB 2312-80, there are vendor character sets developed by Apple for MacOS-S and by IBM for its operating systems. All of these vendor character sets are based on GB 2312-80, as shown in Table C-1.

Table C-1: Vendor Character Set Standards—China

Character Set	Other	User-Defined
DEC Hanzi		8,178
IBM Simplified Chinese	31	1,900 ^a
MacOS-S	4	

^a IBM Simplified Chinese DBCS-Host encoding permits up to 1,900 user-defined characters, but IBM Simplified Chinese DBCS-PC encoding permits only up to 1,880.

DEC Hanzi

The DEC Hanzi character set is the GB 2312-80 character set in the standard 94×94 matrix plus an additional 94×94 matrix set aside for user-defined characters. Specifically, rows 1 through 87 of this additional matrix are for user-defined characters.

IBM Simplified Chinese

The Simplified Chinese character set as used by IBM is based on GB 2312-80, and adds 31 additional characters that are considered to be IBM Simplified Chinese Selected Characters. Depending on the encoding, 1,900 (DBCS-Host) or 1,880 (DBCS-PC) user-defined characters are available.

The 6,763 hanzi in the IBM Simplified Chinese character set are in the same order as in GB 2312-80. The same is true for the GB 2312-80 non-hanzi, but only for DBCS-PC encoding—DBCS-Host encoding uses a completely different ordering for these characters.

The IBM Simplified Chinese DBCS-Host character set, also known as IBM Code Page 00837, is arranged as shown in Table C-2.

Table C-2: IBM Simplified Chinese DBCS-Host Character Set

Character Type	Number of Characters
Full-width space	1
Non-hanzi ^a	712
GB 2312-80 Level 1 hanzi	3,755
GB 2312-80 Level 2 hanzi	3,008
User-defined characters	1,900

^a Includes the 31 IBM Selected Characters.

Table C-3 illustrates the IBM Simplified Chinese DBCS-PC character set, also known as IBM Code Page 00928. Note how the 31 IBM Selected Characters are separate from the GB 2312-80 non-hanzi, but are combined in the IBM DBCS-Host character set.

Table C-3: IBM Simplified Chinese DBCS-PC Character Set

Character Type	Number of Characters
GB 2312-80 non-hanzi	682
GB 2312-80 Level 1 hanzi	3,755
GB 2312-80 Level 2 hanzi	3,008
User-defined characters	1,880
IBM Selected Characters	31

According to older IBM documentation (from 1985), GB 2312-80 79-81 has the form 鍾, but according to more recent IBM documentation (from 1993), the correct form, 钟, is used. The IBM Simplified Chinese DBCS-Host and DBCS-PC encodings for this character are 0x6892 and 0xA891, respectively.

The single-byte portion of IBM's equivalent to GB 1988-89 (GB-Roman) is different from both ASCII and GB-Roman. Table C-4 illustrates the differences among these character sets for two code points.

Table C-4: IBM Variation of GB 1988-89

Code Point	ASCII	GB 1988-89	IBM
24	\$ (dollar)	¥ (yuan)	\$
5C	\ (backslash)	\	¥

The 31 IBM Selected Characters that are included in the IBM Simplified Chinese character set are listed below in a DBCS-PC encoding table.

Row FA	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4		i	ii	iii	iv	v	vi	vii	viii	ix	x					
5	冫	丨	'													
6	一	¥	、	°	、	、										
7	-	全	↗	○	〒	(株)	TEL	··	、	、	▽	▼				
8																
9																
A																
B																
C																
D																
E																
F																

Table C-5 lists the correspondences among these 31 IBM Selected Characters in DBCS-PC and DBCS-Host encodings.

Table C-5: DBCS-PC and DBCS-Host Encoding for 31 IBM Selected Characters

Character	DBCS-PC	DBCS-Host
i	FA41	41B1
ii	FA42	41B2
iii	FA43	41B3
iv	FA44	41B4
v	FA45	41B5
vi	FA46	41B6
vii	FA47	41B7
viii	FA48	41B8
ix	FA49	41B9
x	FA4A	41BA
冫	FA50	425F
丨	FA51	426A
'	FA52	427D
一	FA60	4358
¥	FA61	435B
、	FA62	43BE
°	FA63	43BF

Table C-5: DBCS-PC and DBCS-Host Encoding for 31 IBM Selected Characters (continued)

Character	DBCS-PC	DBCS-Host
\	FA64	43DC
∖	FA65	43DD
-	FA70	445A
仝	FA71	445C
↗	FA72	445E
○	FA73	445F
〒	FA74	446C
(株)	FA75	446D
TEL	FA76	446F
··	FA77	447E
ㄥ	FA78	44DC
ㄨ	FA79	44DD
▽	FA7A	44EB
▼	FA7B	44EC

MacOS-S Character Set

The MacOS-S character set, put simply, is based on the GB 2312-80 character set, encoded according to EUC-CN encoding, with some minor differences in the single-byte range.

The single-byte range, used for encoding ASCII or GB-Roman, uses four additional code points, as illustrated in Table C-6.

Table C-6: Additional MacOS-S Single-Byte Characters

Code Point	Character
80	ü (“u” with diaeresis)
FD	© (copyright)
FE	™ (trademark)
FF	... (ellipsis)

Two of these additional single-byte code points, specifically 0xFD and 0xFE, affect EUC-CN encoding by making rows 0xFD and 0xFE unavailable for encoding two-byte characters (there are no characters in those rows, so there are no adverse effects).

Chinese Vendor Character Sets—Taiwan

Nearly all vendors in Taiwan have standardized on Big Five for their character set, and use it as their base. Thus, most of the vendor character sets described in this section are based on Big Five, as shown in Table C-7.

Table C-7: Vendor Character Set Standards—Taiwan

Character Set	Additional Hanzi	Other	User-defined
DEC Hanyu	13,446 ^a		3,587
ETen	7	399	
IBM Traditional Chinese	3	563	6,204
MacOS-T		34	
Microsoft Traditional Chinese	7	34	

^a CNS 11643-1992 Planes 3 and 4.

DEC Hanyu

The DEC Hanyu (“Hanyu” is the transliterated form of the Chinese word 漢語 *hànyǔ* that means “Chinese”) character set is CNS 11643-1992 Planes 1 through 4 set in four standard 94×94 matrixes. Empty code points in Planes 1 and 2 are used for providing up to 3,587 user-defined characters, specifically the ranges that are provided in Table C-8.

Table C-8: DEC Hanyu User-Defined Regions

Plane	Range	Number of Code Points
1	FDCC–FEFE	145
1	AAA1–C1FE	2,256
2	F245–FE7E	1,186

The most common implementation of DEC Hanyu includes only CNS 11643-1992 Planes 1 and 2 (more or less compatible with Big Five), but Planes 3 and 4 are also available as an extension called DTSCS (Digital Taiwan Supplementary Character Set).

ETen Character Set

ETen (倚天資訊股份有限公司 *yǐtiān zīxùn gǔfēn yǒuxiàn gōngsī*) has developed what appears to be the most widely-used Big Five extension. Microsoft’s Traditional Chinese character set includes the second block of the ETen character set, (that is, the 7 hanzi and 34 non-hanzi at the end of row 0xF9).

The ETen character set's extension to Big Five includes two blocks of characters, as follows:

- The first block that contains 365 characters begins at 0xC6A1 and extends, without gaps, to 0xC8D3. This block includes encircled numerals 1–10, parenthesized numerals 1–10, lowercase Roman numerals 1–10, radicals and radical-like elements, Japanese characters (including kana), the upper- and lowercase Cyrillic alphabet, and ETen input codes (used for ETen's “Row Column” input method^{*}).
- The second block, consisting of seven hanzi (碁锈裏墙恒粧嫫; 0xF9D6 through 0xF9DC) and 34 line-drawing characters, begins at 0xF9D6 and extends, without gaps, to 0xF9FE. Note that 0xF9D5 is the last Big Five hanzi.

The following tables illustrate all 406 ETen-specific characters, as encoded according to Big Five:

Row C6	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4																
5																
6																
7																
A		①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	(1)	(2)	(3)	(4)	(5)
B	(6)	(7)	(8)	(9)	(10)	i	ii	iii	iv	v	vi	vii	viii	ix	x	、
C	ノ	丨	一	冂	冫	彡	勹	匚	卩	ム	夕	宀	ㄩ	彡	彡	彡
D	ヨ	夕	支	无	疒	夂	辵	隶	、	^	、	ゞ	ゞ	ゞ	〃	全
E	々	ゞ	○	一	[]	*	あ	あ	い	い	う	う	え	え	お
F	お	か	が	き	ぎ	く	ぐ	け	げ	こ	ご	さ	ざ	し	じ	
Row C7	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	す	ず	せ	ぜ	そ	ぞ	た	だ	ち	ぢ	つ	づ	つ	づ	て	と
5	ど	な	に	ぬ	ね	の	は	ば	び	ひ	び	び	ふ	ぶ	ぷ	へ
6	べ	べ	ほ	ぼ	ぼ	ま	み	む	め	も	や	や	ゆ	ゆ	よ	よ
7	ら	り	る	れ	ろ	わ	わ	ゐ	ゑ	を	ん	ア	ア	イ	イ	
A		ウ	ウ	エ	エ	オ	オ	カ	ガ	キ	ギ	ク	グ	ケ	ゲ	コ
B	ゴ	サ	ザ	シ	ジ	ス	ズ	セ	ゼ	ソ	ゾ	タ	ダ	チ	ヂ	ツ
C	ツ	ヅ	テ	デ	ト	ド	ナ	ニ	ヌ	ネ	ノ	ハ	バ	パ	ヒ	ビ
D	ピ	フ	ブ	プ	ヘ	ベ	ペ	ホ	ボ	ポ	マ	ミ	ム	メ	モ	ヤ
E	ヤ	ユ	ユ	ヨ	ヨ	ラ	リ	ル	レ	ロ	ワ	ワ	キ	エ	ヲ	ン
F	ヴ	カ	ケ	A	B	B	Г	Д	E	Ё	Ж	З	И	Й	K	

* Written 行列輸入法 (*bángliè shūrùfǎ*) in Chinese.

Row C8	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	Λ	M	H	O	Π	P	C	T	У	Φ	X	Ц	Ч	Ш	Щ	Ъ
5	Ы	Ь	Э	Ю	Я	а	б	в	г	д	е	ё	ж	з	и	й
6	к	л	м	н	о	п	р	с	т	у	ф	х	ц	ч	ш	щ
7	ъ	ы	ь	э	ю	я	↑	↖	⇐	↗	↘	↳	↵	↶	↷	↸
A		ㄣ	ㄨ	ㄩ	ㄚ	1-	58	38	3-	3?	4-	5-	6-	81	7-	8-
B	9-	78	69	9?	01	1?	41	2-	5?	71	48	2?	28	6?	18	18
C	23	33	43	53	63	73	86	93	03	88	93	08	0-	┌	┐	'
D	"	(株)	No.	TEL												
E																
F																

Row F9	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4																
5																
6																
7																
A																
B																
C																
D																
E	┌	┐	└	┘	┙	┚	┛	├	┤	┥	┦	┧	┨	┩	┪	┫
F	┬	┴	┴	┬	┬	┬	┬	┬	┬	┬	┬	┬	┬	┬	┬	┬

It is interesting to note that all seven ETen-specific hanzi map to CNS 11643-1992 Plane 3, as detailed in Table C-9.

Table C-9: Mapping Seven ETen Hanzi to CNS 11643-1992 Plane 3

Hanzi	ETen Code	CNS 11643-1992 Plane 3
碁	F9D6	35-23
鏽	F9D7	47-48
裏	F9D8	36-46
牆	F9D9	48-42
恒	F9DA	12-61
粧	F9DB	29-94
嫺	F9DC	43-60

IBM Traditional Chinese

Table C-10 illustrates the IBM Traditional Chinese DBCS-Host character set, also known as IBM Code Page 00835.

Table C-10: IBM Traditional Chinese DBCS-Host Character Set

Character Type	Number of Characters
Full-width space	1
Non-hanzi	1,003
Big Five Level 1 hanzi	5,402 ^a
Big Five Level 2 hanzi	7,654 ^b
User-defined characters	6,204

^a Includes 1 IBM Traditional Chinese Selected Hanzi appended at the end.

^b Includes 2 IBM Traditional Chinese Selected Hanzi appended at the end.

Table C-11 illustrates the IBM Traditional Chinese DBCS-PC character set, also known as IBM Code Page 00927.

Table C-11: IBM Traditional Chinese DBCS-PC Character Set

Character Type	Number of Characters
Non-hanzi	1,004
Big Five Level 1 hanzi	5,402 ^a
Big Five Level 2 hanzi	7,654 ^b
User-defined characters	6,204

^a Includes 1 IBM Traditional Chinese Selected Hanzi appended at the end.

^b Includes 2 IBM Traditional Chinese Selected Hanzi appended at the end.

The only difference between IBM Traditional Chinese DBCS-Host and DBCS-PC is that the full-width space character is treated specially in DBCS-Host, but included with the other non-hanzi in DBCS-PC. The 1,004 non-hanzi consist of the 441 non-hanzi in Big Five, another 243 that are specific to CNS 11643-1992 (introduced in CNS 11643-1986), and another 320 that are in neither Big Five nor CNS 11643-1992.

As alluded to in the table notes of Tables C-10 and C-11, there are a total of three IBM Traditional Chinese Selected Hanzi. These three hanzi are listed in Table C-12, which also shows how they relate to CNS 11643-1992.

Table C-12: Three IBM Traditional Chinese Selected Hanzi

Hanzi	DBCS-PC	DBCS-Host	CNS 11643-1992
撐	A8CA	68CB	Plane 1 76-93
槩	D1C5	91C6	Plane 2 22-24
濛	D1C6	91C7	Plane 2 36-19

These three IBM Traditional Chinese Selected Hanzi, curiously, are in CNS 11643-1992 Planes 1 or 2. You might be wondering about this, knowing that CNS 11643-1992 Planes 1 and 2 are equivalent to Big Five. Put simply, these three hanzi are considered IBM Selected Hanzi in DBCS-PC and DBCS-Host, and represent the hanzi forms that are now in use by Big Five and CNS 11643-1992. The Level 1 and 2 hanzi in IBM Traditional Chinese are based on hanzi forms found in the original Big Five definition published in 1984. Table C-13 illustrates how these three hanzi and their variant forms are included in IBM Traditional Chinese, and how they map to Big Five and CNS 11643-1992 (those that are considered IBM Traditional Chinese Selected Hanzi—that is, appended at the end of Level 1 and Level 2—are emboldened).

Table C-13: Three IBM Traditional Chinese Selected Hanzi and Their Variants

Hanzi	DBCS-PC	DBCS-Host	CNS 11643-1992	Big Five	Big Five (1984)
撐	A8CA	68CB	Plane 1 76-93	BCB5	
撐	A09C	609D	Plane 3 44-27	F286^a	BCB5
槩	D1C5	91C6	Plane 2 22-24	D5D4	
槩	B3B7	73B8	Plane 3 28-19	F287^a	D5D4
濛	D1C6	91C7	Plane 2 36-19	DE4D	
濛	BAB3	7AB4	Plane 3 34-13	F288^a	DE4D

^a IBM Selected Hanzi code points.

It is interesting that what is considered IBM Selected Hanzi depends on what encoding is used. The older IBM encodings, DBCS-Host and DBCS-PC, share the same definition of what is considered IBM Selected Hanzi, but more contemporary IBM encodings consider the other forms to be IBM Selected Hanzi.

MacOS-T Character Set

The MacOS-T character set is based on the Big Five character set with the addition of 30 annotated numerals (the same 30 that constitute CNS 11643-1992 Plane 1's row 6, specifically encircled, parenthesized, and lowercase Roman numeral forms of 1 through 10).

In addition, the single-byte range, used for encoding ASCII or CNS-Roman, uses four additional code points, as illustrated in Table C-14.

Table C-14: Additional MacOS-T Single-Byte Characters

Code Point	Character
80	ü (“u” with diaeresis)
FD	© (copyright)

Table C-14: Additional MacOS-T Single-Byte Characters (continued)

Code Point	Character
FE	™ (trademark)
FF	... (ellipsis)

Like with the MacOS-S character set, 0xFD and 0xFE are used as single-byte characters, so they are not available for encoding two-byte characters. This affects any vendor extension that encodes characters in either of these two rows. DynaLab's Hong Kong extension, for example, normally encodes characters in rows 0xFD and 0xFE.

Microsoft Traditional Chinese

The Microsoft Traditional Chinese character set, known as Microsoft Code Page 950, is based on Big Five, and includes only the second block of ETen characters, specifically those hanzi and non-hanzi from row 0xF9 (0xF9D6 through 0xF9FE).

Chinese Vendor Character Sets—Hong Kong

While the largest Hong Kong character set is clearly that designed by the Hong Kong government, two major Big Five extensions for Hong Kong have been developed independently of and conflicting with one another. Table C-15 lists three Hong Kong extensions.

Table C-15: Vendor Character Set Standards—Hong Kong

Character Set	Base Set	Hanzi	Other
DynaLab HK A	Big Five	784	
DynaLab HK B	Big Five	665	746
Monotype Hong Kong	Big Five	471	

For more detailed information on these and other Hong Kong extensions, I highly recommend reading Dirk Meyer's informative article entitled "Dealing With Hong Kong Specific Characters" (*MultiLingual Communications & Technology*, Number 19, Volume 9, Issue 3, April 1998, pp 35–38).

DynaLab Hong Kong Extensions

DynaLab (華康科技開發股份有限公司 *huákāng kējì kāifā gǔfēn yǒuxiàn gōngsī*), headquartered in Taiwan, developed two of its own Hong Kong extensions for Big Five. One extension contains 784 hanzi (Dirk Meyer calls this "DynaLab HK A") spread over five rows: 0xFA through 0xFE; and the other contains 1,411 characters (665 hanzi plus 746 symbols, which Dirk calls "DynaLab HK B"). However,

DynaLab's font products for Hong Kong appear to be shifting toward Hong Kong GCCS, which was described in Chapter 3.

The DynaLab HK A set includes several duplicate hanzi. Both groups of duplicate hanzi, listed in Table C-16, consists of 30 hanzi each. The first group, on the left, are those that are dublicately encoded within the DynaLab HK A set itself, and the second group, on the right, are those that are duplicates of standard Big Five hanzi.

Table C-16: Duplicate Hanzi in DynaLab HK A

Hanzi	DynaLab HK A	Hanzi	DynaLab HK A	Big Five
礮	FB52, FD5A	靚	FAC8	E8B0
祇	FB53, FD5E	祇	FB53	ACE9
襖	FB54, FD60	蟄	FB5C	EE68
籽	FB55, FD62	呶	FBD1	A94C
瘿	FB56, FD6E	唉	FBD3	ADFC
痲	FB57, FD70	姘	FBF4	A971
綺	FB59, FDB0	姘	FBF7	ABB9
躑	FB5A, FDB7	帙	FC4A	CC6F
勝	FB5B, FDC5	徼	FC5D	E975
蟄	FB5C, FDD0	悛	FC61	D1AA
蟾	FB5D, FDD3	惚	FC63	B1AB
衤	FB5E, FDDC	憩	FC6F	BECD
袜	FB5F, FDDE	擗	FCA3	BCB5
罨	FB60, FDE5	攢	FCA5	F6E3
莽	FB61, FDEF	朶	FCAC	A6B7
薑	FB62, FDFA	杲	FCB0	AA58
鯢	FB63, FE40	涅	FCD1	AF49
卹	FB64, FE5C	煨	FCE5	B7DA
鉢	FB65, FE70	燁	FCE8	E667
鉢	FB66, FE71	珽	FCFA	D670
颺	FB67, FEB9	祇	FD5E	ACE9
館	FB68, FEBC	祕	FD5F	AFA6
餽	FB69, FEC0	筋	FD77	B5AC
饜	FB6A, FEC2	籙	FDA2	F6FC
駝	FB6B, FEC6	腓	FDC2	B5CC
臚	FB6C, FECA	蟄	FDD0	EE68
髯	FB6D, FECF	祛	FDDF	D7B6
鷄	FB6E, FEEB	讖	FE4A	F9A9

Row FC	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	次	峯	峩	歲	嵒	蔚	嶠	卮	昏	侷	帙	幫	幪	弋	曠	宄
5	灾	弃	寃	窻	廊	迫	园	国	衽	倅	健	徧	僥	徼	悉	恣
6	息	悛	惧	惚	惠	惩	忽	恂	惛	愧	憑	惓	慵	慙	傷	憩
7	憾	慤	慤	慤	慤	威	戾	切	拘	扞	柄	挖	搥	弄	拊	
A		棲	携	携	攢	攢	咬	皓	映	晋	朞	条	朵	析	杵	枘
B	杲	柀	柀	柀	柀	柀	椀	椀	棊	梢	煤	楸	榆	榑	楹	稟
C	桢	桢	棟	槩	欄	檉	欄	檉	檉	欄	鬱	敵	逾	汨	洫	沓
D	沚	涅	湏	湏	湏	潜	濇	濇	濇	濇	濇	焗	焗	焗	焗	焗
E	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗
F	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗	焗

Row FD	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	舄	舄	舄	舄	舄	舄	舄	舄	舄	舄	舄	舄	舄	舄	舄	舄
5	盖	眠	畔	眦	暹	暖	砗	砗	砗	砗	砗	砗	砗	砗	砗	砗
6	襪	襪	籽	杭	球	稽	稟	穉	耗	竝	竝	竝	竝	竝	竝	竝
7	癩	癩	癩	癩	癩	癩	癩	癩	癩	癩	癩	癩	癩	癩	癩	癩
A		纂	纂	纂	纂	纂	纂	纂	纂	纂	纂	纂	纂	纂	纂	纂
B	綺	綳	絳	縵	縵	縵	縵	縵	縵	縵	縵	縵	縵	縵	縵	縵
C	脍	脍	脍	脍	脍	脍	脍	脍	脍	脍	脍	脍	脍	脍	脍	脍
D	蟄	蟄	蟄	蟄	蟄	蟄	蟄	蟄	蟄	蟄	蟄	蟄	蟄	蟄	蟄	蟄
E	衽	衽	衽	衽	衽	衽	衽	衽	衽	衽	衽	衽	衽	衽	衽	衽
F	萌	亂	修	蔥	葱	葱	葱	葱	葱	葱	葱	葱	葱	葱	葱	葱

Row FE	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	覘	覘	覘	詢	詢	詢	詢	詢	詢	詢	詢	詢	詢	詢	詢	詢
5	趨	趨	趨	蹤	蹤	蹤	蹤	蹤	蹤	蹤	蹤	蹤	蹤	蹤	蹤	蹤
6	酖	酖	酖	酖	酖	酖	酖	酖	酖	酖	酖	酖	酖	酖	酖	酖
7	鉢	鉢	鉢	鉢	鉢	鉢	鉢	鉢	鉢	鉢	鉢	鉢	鉢	鉢	鉢	鉢
A		隄	隄	隄	隄	隄	隄	隄	隄	隄	隄	隄	隄	隄	隄	隄
B	韵	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻
C	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻	韻
D	鬃	鬃	鬃	鬃	鬃	鬃	鬃	鬃	鬃	鬃	鬃	鬃	鬃	鬃	鬃	鬃
E	鰐	鰐	鰐	鰐	鰐	鰐	鰐	鰐	鰐	鰐	鰐	鰐	鰐	鰐	鰐	鰐
F	麩	麩	麩	麩	麩	麩	麩	麩	麩	麩	麩	麩	麩	麩	麩	麩

When DynaLab HK A is implemented on MacOS-T, rows 0xFD and 0xFE are not accessible because those two code points are reserved for single-byte characters, as indicated in Table C-14 on page 563.

Row FC	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	礩	礩	衤	衤	襖	秊	稽	稊	穉	窃	窑	窑	竝	豎	筭	筭
5	筭	筭	糲	粮	糲	糲	糲	糲	糲	綫	綉	綉	繚	繚	罇	罇
6	罇	羴	羴	羴	羴	卧	耻	舩	芑	芑	芑	菓	菓	萘	莖	菲
7	葱	蕤	蕤	虬	虬	虬	虬	虬	虬	虬	虬	虬	虬	衞	衞	
A		袜	袴	衤	衤	襖	襖	襖	襖	覩	覩	誦	誦	諷	諷	諷
B	脏	躅	踪	躅	躅	躅	躅	躅	躅	軀	軀	軀	軀	迹	迹	迹
C	邨	酖	酖	酖	酖	酖	鈎	鈎	鈎	鈎	鈎	鈎	鈎	鈎	鈎	鈎
D	鑪	駟	閔	隣	鞞	鞞	韻	韻	颯	飴	餵	餵	餵	餵	餵	餵
E	罵	駟	駟	駟	駟	鬚	鬚	鬚	鬚	鮎	鮎	鮎	鮎	鮎	鮎	鮎
F	鯪	鱸	鱸	鱸	鷓	鷓	鷓	鷓	鷓	麩	麩	麩	麩	麩	麩	麩

The ordering of Monotype's Hong Kong extension follows indexing radical, then total number of strokes. The only exception appears to be the very last hanzi, encoded at 0xFCFE (教).

This Big Five extension is not limited to Monotype's own font products. Arphic Technology, for example, has standardized on Monotype's Hong Kong Big Five extension for their Chinese fonts intended to be used in the Hong Kong market.

Japanese Vendor Character Sets

A large number of Japanese developers have established their own extensions to JIS X 0208:1997 (or earlier versions), usually to accommodate non-JIS characters that are important for their users. The following sections describe many vendor-defined Japanese character sets, all of which are based on JIS X 0208:1997 or earlier. Table C-17 summarizes these character sets.

Table C-17: Vendor Character Set Standards—Japan

Character Set	Base Set	Additional Kanji	Other	User-Defined
Biblos Gaiji Set	<i>not applicable</i>	365	1,406	
DEC Kanji	JIS X 0208:1997			2,914
Enfour Gaiji Set	<i>not applicable</i>	348	1,156	
FMR Kanji	JIS X 0208-1983		3	2,444
Fontworks Japanese	JIS X 0208-1990	530	207	
HP Kanji	JIS X 0208-1983			5,366
IBM Japanese	JIS X 0208-1990	360	28	4,370 ^a
IKIS	JIS X 0208-1983 ^b		63	376
JEF	JIS C 6226-1978	4,039	1,010	3,102 ^c
KanjiTalk6	JIS X 0208-1983		135	2,444
KanjiTalk7	JIS X 0208-1990		313	2,444

Table C-17: Vendor Character Set Standards—Japan (continued)

Character Set	Base Set	Additional Kanji	Other	User-Defined
KEIS78	JIS C 6226-1978	2,042	1,021	3,008
KEIS83	JIS X 0208-1983	2,200	966	3,008
NEC Kanji	JIS C 6226-1978	3,382	1,090	2,256 ^d
NTT Kanji	JIS C 6226-1978	5,238	261	2,820
Super DEC Kanji	JIS X 0208:1997	5,801 ^e	266 ^f	11,374
TRON Code ^g	JIS X 0208-1990	5,801 ^e	266 ^f	27,720
Windows Japanese	JIS X 0208-1990	720 ^h	42	1,880

^a IBM Japanese DBCS-Host encoding permits up to 4,370 user-defined characters, but IBM Japanese DBCS-PC encoding permits only up to 1,880, and IBM Japanese DBCS-EUC encoding permits up to 2,538.

^b That is, JIS X 0208-1983 less 32 line-drawing elements.

^c 457 of these have been pre-assigned under some implementations.

^d There are 2,256 user-defined characters in the Shift-JIS–encoded version of this character set. The ISO-2022-JP–encoded version is usually limited to 188 user-defined characters.

^e JIS X 0212-1990 kanji.

^f JIS X 0212-1990 non-kanji.

^g Can also include GB 2312-80 and KS X 1001:1992.

^h The 360 IBM Selected Kanji appear twice, which is where this figure comes from.

Information about additional Japanese vendor character set standards is available online.*

Biblos Gaiji Character Set

In addition to there being various vendor-specific extensions to JIS X 0208:1997, a company called DTP center Biblos has developed a series of “Gaiji” fonts that have become extremely popular in Japan.[†] One reason for their popularity, especially in professional publishing, is that they match most of Morisawa’s typeface designs. As of this writing, there are Biblos Gaiji sets that match 23 Morisawa typeface designs.

The following tables represent the standard Biblos Gaiji character set, which extends from Shift-JIS row 0xF0 to 0xF9. Rows 0xF0 through 0xF7 contain 1,395 symbols, and rows 0xF8 and 0xF9 contain 365 kanji plus 11 parenthesized kanji numerals. The range 0xF790 through 0xF7FC (109 code points) is left open for user-defined characters.

* <http://www.opengroup.or.jp/jvc/cde/sjis-e.html>

† <http://www.biblosfont.co.jp/>

Row F0	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮	
5	⑰	⑱	⑲	⑳	㉑	㉒	㉓	㉔	㉕	㉖	㉗	㉘	㉙	㉚	㉛	㉜
6	㉝	㉞	㉟	㊱	㊲	㊳	㊴	㊵	㊶	㊷	㊸	㊹	㊺	㊻	㊼	㊽
7	㊾	㊿	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭
8	⑰	⑱	⑲	⑳	㉑	㉒	㉓	㉔	㉕	㉖	㉗	㉘	㉙	㉚	㉛	㉜
9	㉝	㉞	㉟	㊱	㊲	㊳	㊴	㊵	㊶	㊷	㊸	㊹	㊺	㊻	㊼	㊽
A	①	②	③	④	⑤	⑥	a	b	c	d	e	f	g	h	i	j
B	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
C	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
D	Q	R	S	T	U	V	W	X	Y	Z	0	1	2	3	4	5
E	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
F	22	23	24	25	26	27	28	29	30	31	32	33	34			

Row F1	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
5	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66
6	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82
7	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	
8	98	99	100	a	b	c	d	e	f	g	h	i	j	k	l	m
9	n	o	p	q	r	s	t	u	v	w	x	y	z	A	B	C
A	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
B	T	U	V	W	X	Y	Z	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
C	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)
D	(25)	(26)	(27)	(28)	(29)	(30)	(31)	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)
E	(41)	(42)	(43)	(44)	(45)	(46)	(47)	(48)	(49)	(50)	(51)	(52)	(53)	(54)	(55)	(56)
F	(57)	(58)	(59)	(60)	(61)	(62)	(63)	(64)	(65)	(66)	(67)	(68)	(69)			

Row F2	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	(70)	(71)	(72)	(73)	(74)	(75)	(76)	(77)	(78)	(79)	(80)	(81)	(82)	(83)	(84)	(85)
5	(86)	(87)	(88)	(89)	(90)	(91)	(92)	(93)	(94)	(95)	(96)	(97)	(98)	(99)	(100)	(a)
6	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)	(n)	(o)	(p)	(q)
7	(r)	(s)	(t)	(u)	(v)	(w)	(x)	(y)	(z)	(A)	(B)	(C)	(D)	(E)	(F)	
8	(G)	(H)	(I)	(J)	(K)	(L)	(M)	(N)	(O)	(P)	(Q)	(R)	(S)	(T)	(U)	(V)
9	(W)	(X)	(Y)	(Z)	0	1	2	3	4	5	6	7	8	9	10	11
A	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
B	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
C	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
D	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75
E	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
F	92	93	94	95	96	97	98	99	100	a	b	c	d			

Row F3	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t
5	u	v	w	x	y	z	A	B	C	D	E	F	G	H	I	J
6	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
7	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
8	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
9	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46
A	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62
B	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78
C	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
D	95	96	97	98	99	100	a	b	c	d	e	f	g	h	i	j
E	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
F	A	B	C	D	E	F	G	H	I	J	K	L	M			

Row F4	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	0	1	2
5	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
6	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
7	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	
8	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65
9	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81
A	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97
B	98	99	100	a	b	c	d	e	f	g	h	i	j	k	l	m
C	n	o	p	q	r	s	t	u	v	w	x	y	z	A	B	C
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
E	T	U	V	W	X	Y	Z	0.	1.	2.	3.	4.	5.	6.	7.	8.
F	9.	0	1	2	3	4	5	6	7	8	9	○	□			

Row F5	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎
5	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎	☎
6	○	○	●	○	□	▨	▨	◊	◊	◊	◻	◻	◻	◻	◻	◻
7	☐	⇒	⇐	↑	↓	←	⇒	⇑	⇓	⇔	⇕	⇖	⇗	⇘	⇙	⇚
8	←	→	↔	↕	↖	↗	↘	↙	↚	↛	↜	↝	↞	↠	↡	↢
9	▶	◀	↔	↕	↖	↗	↘	↙	↚	↛	↜	↝	↞	↠	↡	↢
A	♠	♠	♠	♠	♠	♠	♠	♠	♠	♠	♠	♠	♠	♠	♠	♠
B	~															I
C	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	i	ii	iii	iv	v
D	vi	vii	viii	ix	x	xi	xii	nm	µm	mm	cm	m	km	mm ²	cm ²	m ²
E	km ²	mm ³	cm ³	m ³	km ³	µg	mg	g	kg	ml	dl	l	kl	ns	µs	ms
F	sec	min	cal	kcal	cc	°C	°F	°K	dB	Hz	ℓ	ℓ	ℓ			

Row F6	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	ミ	セン	キ	メ	ギ	テ	ル	シ	グ	ト	ル	如	ズ	ツ	ル	ズ
5	フ	ビ	バ	ボ	アル	ル	チ	フ	ル	ド	カ	オ	ズ	カ	ル	ル
6	ゼ	シ	フ	マ	リ	ベ	セ	セ	ジ	ゴ	ナ	コ	リ	チ	キ	メ
7	ガ	テ	ル	ロ	ム	ト	ル	ワ	ル	ル	ル	ア	ワ	ビ	バ	
8	イ	ル	ル	チ	ト	ル	ヤ	ヤ	ス	ズ	ゴ	バ	ド	セ	シ	ラ
9	ク	ラ	バ	ベ	キ	ジ	(株)	(有)	(資)	(名)	(相)	(社)	(財)	(学)	(宗)	(医)
A	(協)	(共)	(特)	(監)	(労)	(企)	(機)	(本)	(支)	(営)	(事)	(販)	(研)	(自)	(至)	(普)
B	(当)	(代)	(呼)	(休)	(祝)	(祭)	(日)	(月)	(火)	(水)	(木)	(金)	(土)	()	株式	株式
C	有限	有限	社有	社有	財有	財有	学校	学校	宗教	宗教	印	印	秘	営	問	円
D	控	当	メ	休	祝	祭	日	月	火	水	木	金	土	日	月	火
E	水	木	金	土	印	秘	営	問	円	控	当	メ	休	祝	祭	日
F	月	火	水	木	金	土	日	月	火	水	木	金	土			

Row F7	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	「	」	「	」	「	」	「	」	」	」	」	」	」	」	」	」
5	『	』	『	』	『	』	『	』	』	』	』	』	』	』	』	』
6	”	”	”	”	”	”	”	”	”	”	”	”	”	”	”	”
7	（	）	（	）	〔	〕	〔	〕	〔	〕	〔	〕	〔	〕	〔	〕
8	）	（	）	（	〕	〔	〕	〔	〕	〔	〕	〔	〕	〔	〕	〕
9																
A																
B																
C																
D																
E																
F																

Row F8	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	啞	逢	芦	飴	溢	茨	鰯	淫	迂	嘘	鬱	厩	厩	尊	餌	益
5	焰	襖	鷗	迦	牙	恢	拐	晦	慨	概	喝	葛	鞆	嚙	澗	翰
6	館	翫	徽	祇	吉	汲	笈	俠	卿	教	橋	橋	橋	鄉	饗	僅
7	謹	軀	喰	櫛	屑	靴	乘	乘	薰	祁	惠	慧	稽	繫	荆	
8	隙	倦	嫌	拳	捲	鹵	諺	巷	廣	昂	溝	高	麴	鵠	黑	甌
9	采	冴	榭	崎	寄	柵	薩	鯖	捌	鑄	珊	飼	屢	社	遮	杓
A	灼	爵	繡	酋	遵	曙	渚	緒	薯	諸	諸	哨	廠	鼻	杓	梢
B	祥	蔣	醬	鞘	埴	蝕	侵	浸	真	神	靱	靱	靱	靱	逗	翠
C	摺	瀨	逝	青	靜	靜	蟬	撰	栓	煎	煽	煽	穿	箭	詮	噌
D	遡	創	搔	曹	巢	瘦	揃	遜	驛	腿	黛	臺	才	啄	濯	琢
E	蛸	巽	辿	棚	鱈	樽	簞	註	瀦	猪	凋	抄	槌	鎚	塚	摑
F	辻	鄭	擢	溺	填	顛	堵	屠	菟	賭	土	塘	禱			

Row F9	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	鴛	德	瀆	澗	噸	遁	頓	那	謎	灘	檣	迓	禰	祢	囊	牌
5	這	稗	剝	箸	潑	醞	拔	叛	挽	扉	樋	柶	稗	逼	媛	謬
6	廟	濱	瀕	頻	福	蔽	警	篇	邊	邊	邊	邊	邊	邊	邊	邊
7	婉	庖	泡	蓬	頰	鱒	迄	麵	儲	餅	戾	粃	柳	桺	鑊	
8	愉	愈	癒	猶	猷	猷	祐	熔	耀	萊	賴	隆	龍	遼	漣	煉
9	簾	蓮	朗	榔	蠟	郎	兔	冉	冕	兔	凜	辦	唳	唳	嘲	嚙
A	瑚	媾	宥	屏	悅	振	搆	攢	斃	柎	柎	柎	柎	柎	柎	柎
B	渣	熙	爨	玆	甄	薨	甕	皓	礪	稱	穉	籊	粃	糧	認	縻
C	緝	翊	舩	芍	苒	莫	苾	蔗	蛛	螂	蟒	編	覲	諷	誦	跚
D	跟	輓	迪	邁	夔	霍	靠	頤	鬪	終	鯪	麪	龜	唾	鯪	飴
E	溢	鱗	淫	榘	噉	漚	輝	琪	杵	炫	恍	姤	你	璋	鮭	鄧
F	侖	埆	○	←	←	←	←	←	←	←	←	←	←	←	←	←

There are additional Biblos gaiji products that serve specific purposes, such as characters specifically designed for cloth handling, as illustrated below.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4																
5																
6																
7																

DEC Kanji

Digital Equipment Corporation (デジタルイクイップメント株式会社 *dijitaru ikuippumento kabushikigaisha*) developed two Japanese character set standards: DEC Kanji and Super DEC Kanji.

DEC Kanji consists of a 94×94 matrix identical to that used for JIS X 0208:1997, plus another 94×94 matrix for additional characters. Rows 1 through 31 (2,914 code points) of this additional character space are reserved for user-defined characters, and rows 32 through 94 (5,922 code points) are reserved for private Digital use (but are unused). Rows 9 through 15 and rows 85 through 94 of the JIS X 0208:1997 table are also reserved for Digital use (and are currently unused). Table C-18 shows how characters are allocated to the additional 94×94 matrix.

Table C-18: The DEC Extended Character Set

Rows	Characters	Content
1–31	0	Unassigned (free)
32–94	0	Unassigned (maintained by Digital)

Super DEC Kanji was designed to be a superset of what is available in DEC Kanji and in the full definition of EUC-JP encoding (meaning JIS X 0201-1997, JIS X 0208:1997, and JIS X 0212-1990). It also provides a total of 11,374 user-defined code points, from the encoding regions as illustrated in Table C-19.

Table C-19: Super DEC Kanji User-Defined Regions

Character Set	Rows	Encoding Ranges	Number of Code Points
JIS X 0208:1997	85–94	F5A1–FEFE	940
JIS X 0212-1990	78–94	8FEEA1–8FFEFE	1,598
User-defined	1–94	A121–FE7E	8,836

Enfour Gaiji Character Set

Enfour Media (エヌフォー), like DTP center Biblos, is in the business of developing “Gaiji” products that serve professional publishing needs in Japan. In addition to the standard Enfour Gaiji Set described in this section, Enfour Media also develops Gaiji sets for medical and dental use, and also those that include the IBM and KanjiTalk7 characters (see pages 583 and 588, respectively).

Unlike the Biblos fonts, which match 23 of Morisawa’s typeface designs, Enfour Media’s fonts are considered more generic in design, but do match five of Morisawa’s typeface designs (Mincho Light, Mincho Bold, Gothic Medium, Gothic Bold, and Rounded Gothic Light—try to guess what Morisawa typeface designs these five style and weight combinations correspond to).

The standard Enfour Gaiji Set includes 1,156 non-kanji (most of which are annotated numbers) and 348 kanji (including many JIS78 variants and IBM Selected Kanji), encoded in rows 0xF0 through 0xF7.

The character set tables that span pages 576 through 578 represent the standard Enfour Gaiji character set.

Fontworks Japanese Character Set

Fontworks, a Hong Kong–based Japanese type foundry, provides additional characters beyond JIS X 0208:1997 in their “Pro” series of PostScript font products. This character set includes KanjiTalk7 characters, some JIS78 kanji, some IBM Selected Kanji, and other kanji variants. There are a total of 207 non-kanji and 530 kanji in Fontworks’ extension to JIS X 0208:1997.

The character set tables that span pages 579 and 580 illustrate the characters in Fontworks’ extension to JIS X 0208:1997.

Row F0	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ	Ⓘ	Ⓝ	Ⓚ	Ⓛ	Ⓜ	Ⓝ	Ⓝ	Ⓝ
5	Ⓟ	Ⓠ	Ⓡ	Ⓢ	Ⓣ	Ⓤ	Ⓥ	Ⓦ	Ⓧ	Ⓨ	Ⓩ	+	-	×	÷	=
6	™	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	Ⓟ	Ⓠ	Ⓡ	Ⓢ	Ⓣ	Ⓤ	Ⓥ	Ⓦ	Ⓧ	Ⓨ	Ⓩ	+	-	×	÷	=
8	®	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)	(K)	(L)	(M)	(N)	(O)
9	(P)	(Q)	(R)	(S)	(T)	(U)	(V)	(W)	(X)	(Y)	(Z)	Ā	Ē	Ī	Ō	Ū
A	®	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
B	Ⓟ	Ⓠ	Ⓡ	Ⓢ	Ⓣ	Ⓤ	Ⓥ	Ⓦ	Ⓧ	Ⓨ	Ⓩ	ā	ē	ī	ō	ū
C	Ⓝ	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
D	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51
E	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67
F	68	69	70	71	72	73	74	75	76	77	78	79	80			

Row F1	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96
5	97	98	99	100	(0)	(21)	(22)	(23)	(24)	(25)	(26)	(27)	(28)	(29)	(30)	(31)
6	(32)	(33)	(34)	(35)	(36)	(37)	(38)	(39)	(40)	(41)	(42)	(43)	(44)	(45)	(46)	(47)
7	(48)	(49)	(50)	(51)	(52)	(53)	(54)	(55)	(56)	(57)	(58)	(59)	(60)	(61)	(62)	
8	(63)	(64)	(65)	(66)	(67)	(68)	(69)	(70)	(71)	(72)	(73)	(74)	(75)	(76)	(77)	(78)
9	(79)	(80)	(81)	(82)	(83)	(84)	(85)	(86)	(87)	(88)	(89)	(90)	(91)	(92)	(93)	(94)
A	(95)	(96)	(97)	(98)	(99)	100	0	1	2	3	4	5	6	7	8	9
B	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
C	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41
D	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57
E	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
F	74	75	76	77	78	79	80	81	82	83	84	85	86			

Row F2	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	87	88	89	90	91	92	93	94	95	96	97	98	99	100	0	10
5	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
6	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
7	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	
8	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
9	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89
A	90	91	92	93	94	95	96	97	98	99	100	+	-	×	÷	=
B	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	¢	∠	MB	GB	TB
C	cm	mm ²	mm ³	km ³	l	kl	sec	min	hr	cal	kcal	dB	fff	h	vs.	⌘
D	⌂	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15
E	caps	control	delete	enter	esc	option	return	space	tab	.	<	⌘	⌘	⌘	⌘	⌘
F	🍏	🍏	↔	↗	↖	↘	↙	↗	▶	◀	▷	◁				

Row F3	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
4	☀	☁	☂	☃	☄	○	⊙	◎	●	⦿	⦿	⦿	⦿	⊗	⊗	⊗	
5	▲	●	◎	⊗	⊕	⊗	⊕	⊕	⊕	☃	⊗	⊗	⊗	⊗	⊗	⊗	
6	⌘	⌘	+	☀	☀	⚓	▶	▲	✈	🚗	🚗	🚗	🚗	🚗	🚗	🚗	
7	AND	OR	XOR	NOT	ON	OFF	1/2	1/3	2/3	1/4	3/4	1/5	2/5	3/5	4/5		
8	●	○	■	□	☞	☞	☞	☞	✂	✂	✂	✂	✂	✂	✂	✂	
9	い	ろ	は	に	ほ	へ	と	ち	り	ぬ	る	を	わ	か	よ	た	
A	れ	そ	つ	ね	な	ら	む	う	る	の	お	く	や	ま	け	ふ	
B	こ	え	て	あ	さ	き	ゆ	め	み	し	ゑ	ひ	も	せ	す	ん	
C	い	ロ	ハ	ニ	ホ	ヘ	ト	チ	リ	ヌ	ル	ヲ	ワ	カ	ヨ	タ	
D	レ	ソ	ツ	ネ	ナ	ラ	ム	ウ	ル	ノ	オ	ク	ヤ	マ	ケ	フ	
E	コ	エ	テ	ア	サ	キ	ユ	メ	ミ	シ	エ	ヒ	モ	セ	ス	ン	
F	い	ろ	は	に	ほ	へ	と	ち	り	ぬ	る	を	わ				

Row F4	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
4	(か)	(よ)	(た)	(れ)	(そ)	(つ)	(ね)	(な)	(ら)	(む)	(う)	(る)	(の)	(お)	(く)	(や)	
5	(ま)	(け)	(ふ)	(こ)	(え)	(て)	(あ)	(さ)	(き)	(ゆ)	(め)	(み)	(し)	(ゑ)	(ひ)	(も)	
6	(せ)	(す)	(ん)	(い)	(ロ)	(ハ)	(ニ)	(ホ)	(ヘ)	(ト)	(チ)	(リ)	(ヌ)	(ル)	(ヲ)	(ワ)	
7	(カ)	(ヨ)	(タ)	(レ)	(ソ)	(ツ)	(ネ)	(ナ)	(ラ)	(ム)	(ウ)	(ル)	(ノ)	(オ)	(ク)		
8	(ヤ)	(マ)	(ケ)	(フ)	(コ)	(エ)	(テ)	(ア)	(サ)	(キ)	(ユ)	(メ)	(ミ)	(シ)	(エ)	(ヒ)	
9	(モ)	(セ)	(ス)	(ン)	<small>ミリ</small>	<small>ミリ</small>	<small>メートル</small>	<small>キロ</small>	<small>フィート</small>	<small>アル</small>	<small>マイル</small>	<small>エーカー</small>	<small>ガロン</small>	<small>バレル</small>	<small>ダース</small>	<small>カン</small>	
A	ケース	ノット	ビット	バイト	メガ	ギガ	サイル	ボルト	アンペア	ガウス	ポンド	マルク	フラン	ペセタ	ペソ	リラ	
B	ウソ	ルン	パツ	マン	船	社団法人	学校法人	医療法人	宗教法人	株式会社	チャイ	トイ	ダミ	ダキ	ツル	クサイ	
C	トボル	デン	日	月	火	水	木	金	土	祝	休	祭	呼	内	自	至	
D	企	協	資	社	名	特	注	障	株	有	支	販	営	事	学	普	
E	高	昼	夜	前	後	休	内	注	大	中	小	上	下	左	右		
F	(支)	(販)	(営)	(事)	(普)	(研)	(昼)	(夜)	(前)	(後)	(大)	(中)	(小)	(上)	(下)	(左)	(右)

Row F5	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
4	溢	迂	噂	餌	焰	鷗	迦	恢	晦	喝	葛	嚙	潤	諫	徽	祇	
5	既	廐	俠	卿	軀	櫛	屑	祁	倦	嫌	捲	巷	麴	鵠	甌	采	
6	榭	薩	鯖	鑄	珊	屢	社	繡	酋	曙	渚	薯	諧	梢	蔣	醬	
7	鞘	蝕	逝	蟬	撰	噲	遡	搔	遜	驛	腿	黛	啄	琢	迪		
8	樽	箆	偷	瀦	槌	鎚	塚	擱	壺	鄭	迪	填	堵	屠	賭	禱	
9	瀆	德	滯	遁	謎	檣	禰	囊	蠅	剝	箸	潑	醜	槌	逼	蓬	
A	頰	鱒	麵	儲	餅	粿	鎚	愈	瘡	猷	萊	遼	漣	煉	蓮	榔	
B	蠟	し	丁	几	几	几	几	几	几	几	几	几	几	几	几	几	
C	ヨ	フ	フ	フ	フ	フ	フ	フ	フ	フ	フ	フ	フ	フ	フ	フ	
D	夕	四	内	水	疒	衤	四	主	羽	白	走	食	食	青	麻	黑	
E	D ^b	E ^b	G ^b	A ^b	B ^b	C [#]	D [#]	F [#]	G [#]	A [#]	C ^b	D ^b	E ^b	F ^b	G ^b	A ^b	
F	B ^b	m7	m7	Δ7	+7	(b5)	(#5)	(b9)	(#9)	(b11)	(#11)	(b13)	(#13)				

Row F6	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	shift	clear	help	home	up	down	end	F16	F17	F18	F19	F20	F21	F22	F23	F24
5	☞	←	→	↓	↑	「	」	『	』	⌒	⌒	⌒	⌒	⌒	⌒	⌒
6	ゝ	ゝ	!!	!?	??	!	!!	!?	??	i	い	○	□	□	()	○
7	ゝ	ゝ	//	□	̄	̄	℥	℥	‰	‰	‰	‰	‰	‰	‰	‰
8	μm	μA	μV	μW	μF	μl	∅	ボイ	ヘト	パス	問	答	例	円	男	女
9	本	正	副	写	原	国	公	私	衆	参	相	連	宗	幼	問	答
A	泊	往	復	歩	電	飛	適	項	専	新	旧	再	変	禁	暴	共
B	濟	当	非	税	申	低	朝	本	正	副	写	原	国	公	私	衆
C	参	相	連	宗	幼	泊	往	復	歩	電	飛	専	新	旧	共	当
D	税	控	高	低	朝	医	(一)	(二)	(三)	(四)	(五)	(六)	(七)	(八)	(九)	(十)
E	ルイ	カエ	ンガ	ルイ	スタ	トカ	スケ	トッ	トビ	ガイ	ガメ	ガキ	スガ	ドボ	クル	ンフ
F	タセ	ンベ	ラ	ンウ	フル	ツ	ンボ	トヘ	カス	法社	法学校	法医	法宗	人教		

Row F7	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	啞	飴	或	迤	迤	瑋	昱	今	鰯	淫	英	段	會	會	海	角
5	嵩	神	館	監	寬	熙	熙	濼	崑	起	祺	祈	軌	吉	桴	祛
6	教	高	橋	堯	仰	嶽	柔	柔	慧	患	惠	稽	絜	嶮	鹵	黃
7	廣	廣	昂	昂	高	槁	切	皓	皞	皞	顯	沆	壺	齊	齋	
8	崎	碕	寄	璨	視	你	您	姝	琇	蕓	祝	蔭	緒	翔	祥	祥
9	將	崧	暉	傷	情	穠	丈	耘	埴	真	真	慎	慎	瀨	清	靖
A	精	靜	靜	靜	鮐	箭	操	揃	尊	泰	才	瀧	巽	棚	猪	辻
B	鉄	都	土	唐	鎗	鄧	黨	藤	那	灘	灘	栖	衿	霸	葩	博
C	鏹	扉	彦	濱	敏	樽	福	邊	邊	邊	邊	邊	邊	邊	邊	邊
D	邊	邊	邊	邊	邊	邊	邊	邊	邊	穗	邦	邦	芳	昂	昇	每
E	侖	柳	柳	柳	喻	庾	裕	祐	祐	雄	仔	旺	錫	賴	龍	龍
F	龍	隆	凌	蓼	梁	築	礼	禮	璐	郎	朗	脇	篋			

Fujitsu Character Sets

Fujitsu (富士通 *fujitsu*) has developed two Japanese character set standards, JEF and FMR Kanji. They are used for entirely different environments, specifically Fujitsu's mainframe and personal computers, respectively.

JEF

JEF, short for *Japanese processing Extended Feature*, is the Fujitsu character set primarily used on their mainframe computers called FACOM and some of their OASYS series personal word processors. JEF includes the JIS C 6226-1978 char-

Row F3	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4	蟒	蟬	蠅	蠟	褊	襖	視	觀	角	註	詮	謁	諺	諸	諷	
5	謎	譁	謹	警	謬	賓	賭	贈	趙	跚	跟	蹊	軀	輓	辦	辻
6	迂	迄	迪	迦	迹	迪	道	這	逝	逗	逢	逸	遁	逼	邁	邁
7	遜	遮	邊	遵	邊	邊	邊	邊	邊	邊	邊	邊	那	郎	都	
8	鄉	鄭	鄧	邦	酋	醬	醞	釁	采	釵	鈇	鈇	鑄	錄	鍊	鎚
9	鐘	長	隆	隙	雞	難	霏	雷	青	靜	靜	靠	鞞	鞞	鞞	靴
A	鞞	鞞	響	頤	頤	頤	頤	頤	頤	頤	顛	類	顛	尻	飮	飮
B	飼	餃	餌	餌	餉	餅	餡	餡	饅	饗	饗	騙	騙	驛	高	鬪
C	鬪	鬪	終	鮓	鯁	鯖	鯽	鯽	鯽	鱈	鱈	鵠	鵠	鷗	鶴	鶴
D	麪	麪	麵	麻	黃	黑	墨	黛	鹼	龍	龜	'	.	、	。	、
E	,	'	“	”	”	“	。	,	”		か	け	か	け		
F																

acter set as a subset, along with thousands of other characters. Table C-20 lists the characters included in JEF.

Table C-20: The JEF Character Set—Overview

Character Type	Number of Characters
JIS C 6226-1978	6,802
JEF Extended kanji	4,039
JEF Extended non-kanji	1,010
User-defined characters	3,102 ^a

^a Some implementations of JEF have 457 of these 3,102 code points allocated for kanji specified by Japan's Ministry of Labor (労働省 *rōdōshō*).

As you would expect, the JEF Extended characters are arranged into rows of 94 characters. They are assigned Row-Cell values beginning from 101. Table C-21 shows how JEF Extended characters are allocated to rows 101 through 163.

Table C-21: The JEF Extended Character Set

Rows	Characters	Contents
101–148	4,039	JEF Extended kanji
149–161	917	JEF Extended non-kanji
162	0	Unused
163	93	JEF Extended non-kanji

For the most part, the 4,039 extended kanji are ordered by indexing radical. However, there is what appears to be a compatibility zone containing 71 kanji at the end of the JEF Extended kanji set, which are also ordered by indexing radical.

The JEF Extended kanji area is not fully used—the block of JEF Extended kanji has many empty character positions scattered throughout. JEF actually predates JIS C 6226-1978, and has undergone modifications so that conversion from the JIS X 0208 series is possible.

Most implementations of JEF also include ASCII/JIS-Roman and half-width katakana characters made accessible through the use of shifting characters in an EBCDIC-based encoding.

FMR Kanji

The Fujitsu FM-R series of personal computers make use of a Japanese character set different from JEF both in terms of character set and encoding. This character set is called FMR Kanji. FMR Kanji contains the JIS X 0208 series as its base, yet still makes use of many JIS C 6226-1978 glyphs. It also has three additional hiragana characters not found in JIS X 0208:1997, as illustrated in Table C-22.

Table C-22: Three Non-JIS Hiragana in FMR Kanji

Hiragana	Row-Cell	Transliteration
う	04-84	vu
か	04-85	ka (small version of か)
け	04-86	ke (small version of け)

You may recall from Chapter 3 that there are 83 hiragana and 86 katakana. The difference among these numbers happens to be three characters. The three hiragana characters listed above bring the hiragana set up to 86 characters, like katakana (that is, these three hiragana characters have katakana analogs in JIS X 0208:1997).

Also included in the FMR Kanji character set are the ASCII/JIS-Roman and half-width katakana character set standards, and a user-defined character area that can hold up to 2,444 characters.

Hitachi Character Sets

Hitachi (日立 *hitachi*) developed a character set standard known as KEIS, short for *Kanji processing Extended Information System*. This character set standard comes in two forms: KEIS78 and KEIS83. The former is based heavily on JIS C 6226-1978, and includes 36 additional characters used for print formatting. The latter is based on JIS X 0208-1983. I have a hunch that there is a KEIS90 or KEIS97 in the works or already exists, based on JIS X 0208-1990 or JIS X 0208:1997.

KEIS78

KEIS78 is based on JIS C 6226-1978, and represents Hitachi's original Japanese character set. It also includes 71 non-kanji that were introduced in JIS X 0208-1983. More importantly, its glyphs conform to JIS C 6226-1978, which is important for some classes of users.

KEIS83

After JIS X 0208-1983 was established, Hitachi followed with a version of KEIS that conformed to it. Some shifting of characters took place to accommodate changes that were introduced in JIS X 0208-1983.

Under both KEIS78 and KEIS83, JIS Level 1 kanji and JIS non-kanji are in what Hitachi calls the KEIS Basic Character Set. JIS Level 2 kanji in its entirety makes up KEIS Extended Character Set 1. Corporate-defined kanji and non-kanji are in KEIS Extended Character Set 3 (the kanji are arranged by radical then stroke count, identical to JIS Level 2 kanji). Oddly enough, there is no mention of a KEIS Extended Character Set 2. Table C-23 lists the number of characters in KEIS78 and KEIS83, arranged by its three character sets.

Table C-23: The KEIS78 and KEIS83 Character Sets

	Basic Character Set	Extended Character Set 1	Extended Character Set 3
KEIS78	3,454 ^a	3,384	3,027 ^b
KEIS83	3,489	3,388	3,166 ^c

^a This is the same as JIS C 6226-1978 Level 1 kanji and non-kanji plus 36 formatted printing characters.

^b This includes 71 JIS X 0208-1983 non-kanji, 914 Hitachi non-kanji, and 2,042 kanji.

^c This includes 966 Hitachi non-kanji and 2,200 kanji.

KEIS78 and KEIS83 also includes a user-defined character range. This area can hold up to 3,008 characters (equivalent to 32 rows of 94 characters).

HP Kanji

The Japanese character set implemented by Hewlett-Packard (HP) consists of the 94 printable ASCII/JIS-Roman characters, the 63 half-width katakana characters, and JIS X 0208-1983. Nothing out of the ordinary here. However, in Appendix D, *Vendor Encoding Methods*, which discusses HP Kanji's encoding methods, you will see a departure from this apparent lack of ordinariness in that there is a large user-defined character area that can hold up to 5,366 characters.

IBM Japanese

IBM (アイ・ビー・エム *ai bi emu*) was one of the first companies to develop a Japanese vendor character set standard, called “IBM Japanese” in the scope of this book. This vendor character set standard includes those characters from JIS X 0208:1997 plus an additional 360 kanji and 28 non-kanji known as IBM Selected Kanji and IBM Selected Non-kanji, respectively. IBM Japanese has followed the JIS X 0208 standard very closely. For example, when JIS X 0208-1990 superseded JIS X 0208-1983 in late 1990, IBM quickly moved to standardize to JIS X 0208-1990 by including the two kanji 凜 (84-05) and 熙 (84-06).

IBM Japanese does have a peculiar twist, though. There are three encodings for this character set, and although these encoding methods handle the same set of characters, their characters are arranged differently. Table C-24 shows how the characters are defined under one encoding method, IBM Japanese DBCS-Host, also known as IBM Code Page 00300.

Table C-24: IBM Japanese DBCS-Host Character Set

Character Type	Number of Characters
Full-width space	1
Non-kanji	551
Basic kanji	3,226
Extended kanji	3,489
User-defined characters	4,370

Under the other two encoding methods, IBM DBCS-PC and IBM DBCS-EUC, you clearly see that the break-down of IBM Japanese is quite different. IBM Japanese DBCS-PC is also known as IBM Code Page 00301. Table C-25 illustrates this, and note that the first three entries of Table C-25 comprise the JIS X 0208:1997 character set.

Table C-25: IBM Japanese DBCS-PC and DBCS-EUC Character Set

Character Type	Number of Characters
Non-kanji	524
JIS X 0208 Level 1 kanji	2,965
JIS X 0208 Level 2 kanji	3,390
IBM Selected Non-kanji	28
IBM Selected Kanji	360
User-defined characters	1,880 ^a

^a IBM DBCS-EUC encoding permits up to 2,538 user-defined characters.

With the exception of the number of user-defined characters, the total number of characters is identical, specifically 7,267. The current number of IBM Selected Non-kanji happens to be 26. The difference between that number and the number in the above table, 28, consists of two characters that were not included in JIS C 6226-1978, but became part of JIS X 0208-1983. After IBM standardized to JIS X 0208-1983, these two characters were dropped, thus reducing IBM Selected Non-kanji to 26 characters. I refer to 28 IBM Selected Non-kanji for backward compatibility, and because some implementations include all 28 of these characters, such as Microsoft's Japanese character set (see page 591). These two characters are shown in Table C-26.

Table C-26: Special Mappings for Two IBM DBCS-PC Characters

Character	IBM DBCS-PC Code	JIS X 0208:1997
⌘	FA54	02-44
⌘	FA5B	02-72

I once conducted a short study in which I tried to match the 360 IBM Selected Kanji with kanji from JIS X 0212-1990. The outcome of this study was that 279 kanji in JIS X 0212-1990 matched those in IBM Selected Kanji. There was even one that matched a kanji in JIS X 0208:1997. The remaining 80 kanji map to the JIS X 0212-1990 user-defined region, and in fact there are two such mappings as listed in Appendix Q, *Character Lists and Mapping Tables*. Approximately 70 of these remaining 80 kanji are common to the JEF character set standard, covered earlier in this chapter starting on page 578.

IBM Japanese also includes the ASCII/JIS-Roman and half-width katakana character sets. These fall into what is called SBCS (Single-Byte Character Set). A DBCS (Double-Byte Character Set) and an SBCS (Single-Byte Character Set) together are referred to as an MBCS (Multiple-Byte Character Set).

Other vendors have included IBM Selected Kanji and Non-kanji into their products or even into their own character set standards. As an example, some of NEC's PC-9800 computer systems include all 360 IBM Selected Kanji in rows 89 through 92 of JIS X 0208, and 14 of the 28 IBM Selected Non-kanji in the remainder of row 92 (see page 592 for more details). Microsoft has also included the IBM Selected Kanji and Non-kanji in two regions, specifically in regions defined by *both* IBM and NEC. The character set tables on page 585 illustrate the characters specific to IBM Japanese, specifically IBM Selected Kanji and Non-kanji.

IKIS

Nippon (Japan) Data General (日本データゼネラル *nippon dēta zeneraru*) developed a character set standard very similar to JIS X 0208-1983, except that it

contains the half-width katakana character set within the 94×94 character space—these characters are placed in row 8. In addition, only rows 9 through 12 are assigned as a user-defined character space. This character set standard is referred to as IKIS, which stands for *Interactive Kanji Information System*. Table C-27 illustrates the differences between IKIS and JIS X 0208-1983.

Table C-27: Comparing JIS X 0208-1983 and IKIS

Rows	Characters	Content
8	63	Half-width katakana
9–12	0	Unassigned (free)
13–15	0	Unassigned (reserved)

K-JIS

Developed by 共同通信社 (*kyōdō tsūshinsha*) and 配信先新聞社 (*haishinsen shinbunsha*) for writing newspaper articles. The book entitled 記者ハンドブック (*kisha handobukku*) includes a complete listing of K-JIS-specific kanji and non-kanji. Also of interest is that many K-JIS-specific kanji are part of JIS X 0212-1990.

MacOS-J Character Sets

Apple Computer (アップルコンピュータ *appuru konpyūta*) developed their own Japanese character set with the introduction of KanjiTalk (漢字Talk *kanji tōku*), the Japanese operating system for the Macintosh computer, which is now called MacOS-J. This character set is based on JIS X 0208-1983, but has 82 additional characters in row 13, and 53 vertically-set variants. I call this character set the KanjiTalk6 character set. This character set is implemented in MacOS-J prior to Version 7.1.

The KanjiTalk6 and KanjiTalk7 character sets share the same set of characters for the one-byte range, which is ASCII/JIS-Roman plus four additional characters. These four additional characters are illustrated in Table C-28.

Table C-28: Additional KanjiTalk6 and KanjiTalk7 Single-Byte Characters

Code Point	Character
80	\ (backslash)
FD	© (copyright)
FE	™ (trademark)
FF	... (ellipsis)

Row 14	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
00	あ		い		う		え		お										
20															っ				
40																			
60							や		ゆ		よ								わ
80																			

Row 15	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
00	ア		イ		ウ		エ		オ										
20															ツ				
40																			
60							ヤ		ユ		ヨ								ワ
80					カ	ケ													

Although KanjiTalk6 (the operating system) is now years old and no longer used, the KanjiTalk6 character set is still being used by many Japanese fonts for MacOS-J, including those developed by Adobe Systems and Morisawa.

The KanjiTalk7 character set

Apple Computer developed a new Japanese character set with the introduction of KanjiTalk 7.1 in late 1992. This character set is based on JIS X 0208-1990. I call this Japanese character set the KanjiTalk7 character set. This character set contains ASCII/JIS-Roman, half-width katakana, and JIS X 0208-1990. Rows 9 through 15 of this character set standard contain 260 characters above and beyond JIS X 0208-1990, plus the same 53 vertical variants that are included in the KanjiTalk6 character set. Table C-30 illustrates the differences between JIS X 0208-1990 and the KanjiTalk7 character set.

Table C-30: Comparing JIS X 0208:1997 and KanjiTalk7 Character Sets

Rows	Characters	Content
9	59	Encircled numerals 1–20, parenthesized numerals 1–20, black-encircled numerals 1–9, numerals 0–9 with period
10	56	Upper- and lowercase Roman numerals 1–15, parenthesized lowercase Latin characters
11	34	Abbreviations
12	27	Miscellaneous symbols
13	39	25 parenthesized kanji, 14 encircled kanji
14	35	28 katakana ligatures, 4 two-kanji ligatures, 3 four-kanji ligatures
15	10	5 miscellaneous symbols, 1 hiragana, 4 katakana
85	31	Vertical variants of row 1 (miscellaneous symbols)

There are some characters used in Japanese, mainly punctuation marks, parentheses, and small versions of kana, that need to be positioned differently within their em-square when set vertically. Vertical Japanese text is described in Chapter 7. These characters are found in rows 1, 4, and 5. Rows 85, 88, and 89 in the KanjiTalk7 character set contain the vertical variants of rows 1, 4, and 5, respectively. Likewise, rows 11, 14, and 15 in the KanjiTalk6 character set contain the vertical variants of rows 1, 4, and 5, respectively. This difference between the KanjiTalk6 and KanjiTalk7 character sets is not found in the characters contained in these rows, but rather in the offsets used. Table C-31 details this difference between the KanjiTalk6 and KanjiTalk7 character sets.

Table C-31: Vertical Character Positions in the KanjiTalk6 and KanjiTalk7 Character Sets

	Row Offset	Row 1	Row 4	Row 5
KanjiTalk6	10	11	14	15
KanjiTalk7	84	85	88	89

You can imagine what a headache this row offset value change caused developers who produced software that relied on a value of 10 to access the vertically-set variants of those rows. Apple Computer has plans to eventually phase out these vertically-set variants altogether. I don't mean that you will no longer be able to set Japanese vertically on a Macintosh, but that they will be removed from the character set. This may mean that they are stored internally at the same code positions as their horizontally-set counterparts.

The KanjiTalk6 and KanjiTalk7 character sets both provide 2,444 user-defined character positions. This amounts to 13 rows of 188 characters, which is equivalent to 26 rows of 94 characters.

The PostScript equivalent of the KanjiTalk6 character set is *fontname-83pv-RKSJ-H*, and the PostScript equivalent of the KanjiTalk7 character set is *fontname-90pv-RKSJ-H*. See Chapter 6, *Font Formats*, for more details.

Microsoft Japanese

The character sets in Windows 3.1J and Win95J are identical, and can be described as JIS X 0208-1990 with NEC Row 13 plus the IBM Selected Kanji and Non-kanji sets (in both IBM and NEC positions). See the IBM Japanese and NEC Kanji sections for more information on these character sets, on pages 583 and 592 of this chapter, respectively. For details about Shift-JIS encoding, see page 175 in Chapter 4, *Encoding Methods*. This character set is also known as Microsoft Code Page 932.

The Japanese fonts bundled with Microsoft Windows and some of Microsoft's Japanese applications, such as Microsoft Word-J, were recently expanded to include all of JIS X 0212-1990, along with hundreds of additional characters found in Unicode.

The PostScript equivalent of the Windows 95J character set is *fontname-90ms-RKSJ-H* (horizontal) and *fontname-90ms-RKSJ-V* (vertical). See Chapter 6 for more details.

NEC Kanji

Nippon (Japan) Electronics Corporation (NEC; 日本電気株式会社 *nippon denki kabushikigaisha*) developed its own character set for use on its personal computers and dedicated Japanese word processors. This character set is based on JIS C 6226-1978, and also includes JIS-Roman and half-width katakana. The basic NEC Kanji character set also includes the 360 IBM Selected Kanji and 14 of the 28 IBM Selected Non-kanji set into rows 89 through 92. There are three ways in which the NEC implementation of the IBM Selected Kanji and Non-kanji differ from IBM's own implementation:

- They are encoded in a different region, specifically within the 94×94 matrix
- The 360 kanji come first followed by the non-kanji
- Only 14 of the 28 IBM Selected Non-kanji are included because the remaining 14 characters are already included in NEC Row 13

Table C-32 lists the differences between NEC Kanji and JIS X 0208:1997 (note that row numbers 2, 8, and 84 are identical to JIS C 6226-1978—compare with Table 3-62 on page 106).

Table C-32: The Differences Between JIS X 0208:1997 and NEC Kanji

Rows	Characters	Content
2	14	Miscellaneous symbols
8	0	Unassigned (reserved)
9	94	Half-width JIS-Roman characters
10	94	63 standard half-width katakana, 31 additional half-width katakana
11	93	76 half-width line-drawing elements, 17 half-width miscellaneous symbols
12	76	Full-width line-drawing elements
13 ^a	83	Circled numerals 1–20, uppercase Roman numerals 1–10, 16 katakana ligatures, 10 abbreviations, 4 two-kanji ligatures, 5 encircled kanji, 3 parenthesized kanji, 15 miscellaneous symbols
84	0	Unassigned (reserved)
89–92	374	360 IBM Selected Kanji and 14 IBM Selected Non-kanji

^a This row initially contained only 82 characters. Sometime after 1989 the two-kanji ligature for the Japanese era name 平成 was added to this row to bring the total up to 83. This character is a ligature form of the kanji compound 平成 (*heisei*), which stands for the Heisei Era (1989–present).

Note the inclusion of half-width characters here, specifically the JIS-Roman and half-width katakana character sets. Some implementations include one-fourth size characters in rows 14 and 15.

NEC Kanji can also include an extended character set. Let's call this NEC Extended Kanji. These characters are arranged into a separate 94×94 matrix, and include 682 non-kanji and 3,382 kanji. Table C-33 lists the contents of this extended character set.

Table C-33: The NEC Extended Kanji Character Set

Rows	Characters	Content
1	94	Miscellaneous symbols
2	93	Miscellaneous symbols
3	92	Miscellaneous symbols
4	94	Miscellaneous symbols
5	94	Miscellaneous symbols
6	63	41 katakana ligatures, 20 parenthesized kanji, 2 encircled kanji
7–15	0	Unassigned (free)
16–17	152	Cursive kana characters
18–53	3,382	Kanji arranged by radical then stroke count
54–94	0	Unassigned (free)

Many of these 3,382 kanji are common to JIS X 0212-1990. NEC Extended Kanji was developed nearly 10 years before JIS X 0212-1990, and doesn't appear to be in very common use.

There seems to be a shift at NEC whereby its character set is becoming compatible with JIS X 0208:1997. Two NEC products released in late 1991 (both dedicated Japanese word processors) boasted support for the JIS X 0208-1990 character set.

The following tables illustrate how NEC Row 13, IBM Selected Kanji, and IBM Selected Non-kanji are encoded according to the NEC Kanji character set:

Row 13	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19
00	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮	⑯	⑰	⑱	⑲
20	㊾	I	II	III	IV	V	VI	VII	VIII	IX	X	ミ	キ	セ	トル	グラム	トン	アル	ケル
40	トル	ワ	カ	ドル	セ	セント	ドル	シ	mm	cm	km	mg	kg	cc	m ²				
60			戦	シ	シ	No.	K.K.	TEL	上	中	下	左	右	(株)	(有)	(代)	明治	大正	昭和
80	≡	≡	∫	§	Σ	√	⊥	∠	∟	△	∴	∩	∪						

Row 89	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	
00		續	嬰	鎡	銑	齏	倍	炆	昱	精	銀	昇	彌	丨	仡	任	佂	仔	但	佻
20	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻	佻
40	苧	九	荔	勅	勛	勻	勿	匡	邵	厓	厲	規	夔	吃	味	咩	咭	咭	咭	咭
60	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠
80	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠	垠
Row 90	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	
00		恣	愨	悅	愨	愨	愨	愨	愨	愨	愨	愨	愨	愨	愨	愨	愨	愨	愨	
20	昀	昕	昂	昉	昉	昉	昉	昉	昉	昉	昉	昉	昉	昉	昉	昉	昉	昉	昉	
40	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘	榘
60	汜	沆	沆	沆	沆	沆	沆	沆	沆	沆	沆	沆	沆	沆	沆	沆	沆	沆	沆	沆
80	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆	瀆
Row 91	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	
00		狢	狢	狢	狢	狢	狢	狢	狢	狢	狢	狢	狢	狢	狢	狢	狢	狢	狢	
20	峻	皂	皜	皜	皜	皜	皜	皜	皜	皜	皜	皜	皜	皜	皜	皜	皜	皜	皜	
40	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝	竝
60	藁	蒴	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓	蕓
80	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱	諱
Row 92	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	
00		釗	釗	釗	釗	釗	釗	釗	釗	釗	釗	釗	釗	釗	釗	釗	釗	釗	釗	
20	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	鉸	
40	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛	鑛
60	飯	飼	餒	館	醇	麟	高	謁	魴	魴	魴	魴	魴	魴	魴	魴	魴	魴	魴	
80		i	ii	iii	iv	v	vi	vii	viii	ix	x	一	丨	！	”					

Some of the characters in NEC Row 13 come in slightly different forms depending on the implementation. Table C-34 lists some of these characters, along with a known variation.

Table C-34: Alternate Forms for NEC Row 13 Characters

NEC Form	Alternate Form
セ	セ
グ	グ
アル	アル
ワット	ワット
セント	セント
ページ	ページ

Table C-34: Alternate Forms for NEC Row 13 Characters (continued)

NEC Form	Alternate Form
KK	KK
TEL	TEL

NTT Kanji

Nippon (Japan) Telegraph and Telephone (NTT; 日本電信電話 *nippon denshin denwa*) developed a character set that includes a non-kanji portion identical to JIS X 0208-1983, and a kanji portion identical to JIS C 6226-1978. There are also 261 NTT-specific symbols in the non-kanji region. These include lowercase and uppercase Roman numerals, additional mathematical symbols, symbols for units of measurement, additional line-drawing characters, and graphic representations for ASCII control characters.

NTT Kanji includes an additional 94×94 character space for kanji above and beyond those specified in JIS C 6226-1978. The first 60 rows of this additional character space (5,640 code points) have 5,238 kanji allocated to them, 4,048 of which are kanji found in the dictionary entitled 新字源 (*shinjigen*, meaning “new character origins”), but not in JIS C 6226-1978; the remaining 1,190 kanji are for use in writing person and place names. Within the first 60 rows of this character space, rows 1 through 44 are called Level 1 (4,048 kanji), and rows 45 through 57 are called Level 2 (1,190 kanji). Rows 61 through 64 (376 total code points) are allocated for extended non-kanji, but have yet to be assigned characters. Rows 65 through 94 (2,820 total code points) are reserved for user-defined characters. Table C-35 shows how characters are allocated to the additional 94×94 matrix.

Table C-35: The NTT Kanji Character Set

Rows	Characters	Content
1–44	4,048	Level 1 kanji
44–60	1,190	Level 2 kanji
61–64	0	Unassigned (reserved for extended non-kanji)
65–94	0	Unassigned (free)

TRON Character Set

The TRON character set, which is used on various instances of TRON, such as BTRON, is composed of four zones, labelled A through D, and supports the JIS X 0208:1997 and JIS X 0212-1990 character sets.* JIS X 0208:1997 and JIS X 0212-1990 are allocated to zones A and B, respectively.

* <http://tron.um.u-tokyo.ac.jp/> or <http://www.tokyoweb.or.jp/tron/>

Zone A is used for the most commonly-used character set. For Japanese, this is obviously JIS X 0208:1997. Zone B, as expected, is used for the next most commonly-used character set, which is JIS X 0212-1990. A recent development in TRON is that China's GB 2312-80 and Korea's KS X 1001:1992 are allocated to Zones C and D, respectively.

Korean Vendor Character Sets

There are two important Korean vendor character sets, specifically those for the Korean versions of MacOS and Microsoft Windows, along with IBM's Korean implementations. All are based on KS X 1001:1992, as shown in Table C-36.

Table C-36: Vendor Character Set Standards—Korea

Character Set	Additional Hangul	Additional Hanja	Other	User-Defined
DEC Korean				
HangulTalk			1,137	188
IBM Korean	270	377	6	1,880 ^a
Unified Hangul Code	8,822			

^a IBM Korean DBCS-Host encoding permits up to 1,880 user-defined characters, but IBM Korean DBCS-PC encoding permits only up to 1,227.

DEC Korean

The DEC Korean character set is identical to KS X 1001:1992, and appears to provide no support for user-defined characters.

HangulTalk Character Set

The HangulTalk character set, which is used on MacOS-KH, was originally developed by Elex Computer. It is based on the KS X 1001:1992 character set, but adds 1,137 additional characters, many of which are typeface-independent. Elex Computer designed this character set, and includes many symbols that are normally available in proprietary typesetting systems.

The single-byte range, used for encoding ASCII or KS-Roman, uses five additional code points, as illustrated in Table C-37.

Table C-37: Additional HangulTalk Single-Byte Characters

Code Point	Character
81	₩ (“won” symbol)
82	- (minus)
83	© (copyright)

Table C-37: Additional HangulTalk Single-Byte Characters (continued)

Code Point	Character
FE	™ (trademark)
FF	... (ellipsis)

The following tables illustrate the 1,137 characters that make up the KS X 1001:1992 extension of the HangulTalk character set, arranged according to their encoding:

Row A1	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4		「	」	『	』	『	』	『	』	《	》	<	>	()	〔
5)	(()	(()	(()	(()	(()	【	】	〔	〕	〔
6)	[]	[]	[]	[]	[]	†	*	**	***	”
7	,	+	-	x	∞	∞	±	±	>	<	=	≠	()		
8		{	}	[]	∩	∪	∩	∪	∩	∪	∩	∪	∩	∪	∩
9	五	六	七	八	九	十	●	○	,	[]		*	*	*	*
A	**															
B																
C																
D																
E																
F																

Row A2	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
4		(()	(()	()	〔	〕	【	】	()				
5																	
6																	
7																	
8																	
9																	
A																	
B																	
C																	
D																	
E																	
F	11	12	13	14	15	16	17	18	19	20	¶			:	::		

Row A3	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4		①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮
5	⑯	⑰	⑱	⑲	⑳	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪
6	⑫	⑬	⑭	⑮	⑯	⑰	⑱	⑲	⑳	(A)	(B)	(C)	(D)	(E)	(F)	(G)
7	(H)	(I)	(J)	(K)	(L)	(M)	(N)	(O)	(P)	(Q)	(R)	(S)	(T)	(U)		
8		(V)	(W)	(X)	(Y)	(Z)	(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)	(J)
9	(K)	(L)	(M)	(N)	(O)	(P)	(Q)	(R)	(S)	(T)	(U)	(V)	(W)	(X)	(Y)	(Z)
A																
B																
C																
D																
E																
F																

Row A4	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4		①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮
5	⑯	⑰	⑱	⑲	⑳	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪
6	⑫	⑬	⑭	⑮	⑯	⑰	⑱	⑲	⑳	①	②	③	④	⑤	⑥	⑦
7	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮	⑯	⑰	⑱	⑲	⑳	±		
8		<	>	^	v	†	‡	≠	≠	≈	≈	≡	≡	≡	≡	≡
9	≧	≧	≧	≧	≧	≧	≧	≧	◁	≈	≈	≈	∖	T	≡	≡
A																
B																
C																
D																
E																
F																

Row A5	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4		①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮
5	⑯	⑰	⑱	⑲	⑳	A)	B)	C)	D)	E)	F)	G)	H)	I)	J)	K)
6	L)	M)	N)	O)	P)	Q)	R)	S)	T)	U)	V)	W)	X)	Y)	Z)	a)
7	b)	c)	d)	e)	f)	g)	h)	i)	j)	k)	l)	m)	n)	o)		
8		p)	q)	r)	s)	t)	u)	v)	w)	x)	y)	z)				
9																
A																
B																
C																
D										!	°	'	"	'''		
E																
F										⑳	㉑	㉒	㉓			

Row A9	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4		A.	B.	C.	D.	E.	F.	G.	H.	I.	J.	K.	L.	M.	N.	O.
5	P.	Q.	R.	S.	T.	U.	V.	W.	X.	Y.	Z.	a.	b.	c.	d.	e.
6	f.	g.	h.	i.	j.	k.	l.	m.	n.	o.	p.	q.	r.	s.	t.	u.
7	v.	w.	x.	y.	z.											
8																
9																
A																
B																
C																
D																
E																
F																

Row AA	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4		문	답	주	명	대	형	부	전	접	수	동	비	반	자	타
5	갑	약	인	뜻	印	註	예	感	冠	答	代	頭	動	名	目	反
6	補	本	副	序	連	影	例	源	子	前	節	接	助	指	他	派
7	形	조	문	답	주	뜻	註	교	역	음	정	해	예	존		
8		라	마	바	사	아	자	차	카	타	파	눌	낮	명	대	형
9	부	전	접	수	동	비	계	반	속	인	본	약	속	유	관	冠
A																
B																
C																
D																
E																
F					(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	

Row AB	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4		조	국	갑	印	衣	未	거	답	변	상	센	신	여	예	원
5	작	준	큰	외	활	간	갈	실	感	慣	代	動	名	反	副	自
6	前	電	接	助	注	參	本	新	現	形	間	國	中	他	빠	시
7	입	으	음	직	표	가	나	다	하	마	바	사	아	자		
8		차	카	타	파	하	비	답	빠	본	단	센	시	여	예	으
9	음	입	제	존	준	표	해	느	눌	낮	반	가	나	다	라	의
A																
B																
C																
D																
E																
F								(27)	(28)	(29)	(30)					

Row AC	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4		↗	↖	↘	↙	↔	↞	↠	↡	↢	↣	↤	↥	↦	↧	↨
5	↩	↪	↫	↬	↭	↮	↯	↰	↱	↲	↳	↴	↵	↶	↷	↸
6	↙	↘	↗	↖	↕	↔	↞	↠	↡	↢	↣	↤	↥	↦	↧	↨
7	↗	↖	↘	↙	↕	↔	↞	↠	↡	↢	↣	↤	↥	↦	↧	↨
8		↩	↪	↫	↬	↭	↮	↯	↰	↱	↲	↳	↴	↵	↶	↷
9	☞	단	참	소	중	일	이	휴								
A																
B																
C			1	2	3	4	5	6	7	8	9	10	11	12	13	14
D	15															
E																
F			16	17	18	19	20									

Row AD	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
4		一	二	三	四	五	六	七	八	九	十	士	吉	吉	吉	吉
5	夫	老	大	夫	早	一	二	三	四	五	六	七	八	九	十	士
6	吉	吉	吉	吉	夫	老	大	夫	早	日	月	火	水	木	金	士
7	日	月	火	水	木	金	士	日	月	火	水	木	金	士		
8																
9																
A		「	」	『	』	⇒	⇐	↑	↓	”	„	˘	˙	˚	˛	↑
B	!															
C																
D																
E																
F																

I have noticed that some Korean type foundries have the Yin-Yang (음양/陰陽 *eumyang*, meaning “negative [and] positive”) symbols reversed when implementing this character set. This symbol is called 태극/太極 (*taegeug*) in Korean. Row 0xA6 includes three instances of this character: ☯ (0xA693), ☯ (0xA697), and ☯ (0xA698). Notice how they appear to be rotating in the counter-clockwise direction, which is considered to be correct.

Many of the gaps in the HangulTalk character set are actually by design, specifically so that it can be overlaid on the KS X 1001:1992 character set encoded according to EUC-KR with no overlapping of characters. You can compare these tables with the KS X 1001:1992 code table in Appendix L, *KS X 1001:1992 Table*, to see how they overlay.

IBM Korean

The IBM Korean character set is based on KS X 1001:1992, and adds 377 hanja, 270 hangul, and 6 symbols. As with other IBM character sets, their Korean character set has been implemented using two different encodings: DBCS-Host and DBCS-PC.

The IBM Korean DBCS-Host character set, also known as IBM Code Page 00834, is arranged as shown in Table C-38.

Table C-38: IBM Korean DBCS-Host Character Set

Character Type	Number of Characters
Full-width space	1
Symbols ^a	991
Hanja ^b	5,265
Hangul ^c	2,620
User-defined characters	1,880

^a Includes the 6 IBM Selected Characters.

^b Includes the 377 IBM Selected Hanja

^c Includes the 270 IBM Selected Hangul

Interestingly, the 2,620 hangul that are included in IBM Korean DBCS-Host are encoded according to Johab encoding, as described in Chapter 4 starting on page 177.

Table C-39 illustrates the IBM Korean DBCS-PC character set, also known as IBM Code Pages 00926 or 00951. Note how the three sets of IBM Selected Characters are separate from the KS X 1001:1992 characters, but are combined with them in the IBM DBCS-Host character set. Also note how the 1,227 user-defined characters are separated into three regions. The last two user-defined regions, with 94 available code points each, are from KS X 1001:1992 itself.

Table C-39: IBM Korean DBCS-PC Character Set

Character Type	Number of Characters
User-defined characters	1,039
IBM Selected Characters	6
IBM Selected Hanja	377
IBM Selected Hangul	270
KS X 1001:1992 symbols	986
KS X 1001:1992 hangul	2,350
User-defined characters	94

Table C-39: IBM Korean DBCS-PC Character Set (continued)

Character Type	Number of Characters
KS X 1001:1992 hanja	4,888
User-defined characters	94

The following tables illustrate the IBM Korean Selected Characters, encoded according to IBM Korean DBCS-PC encoding (where these IBM-specific characters are consolidated to seven rows):

Row 9A	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A							≠	≤	≥	°K	！	”	咯	攔	羯	酣
B	釀	骼	蒹	哽	扃	榮	頰	畊	稷	雞	杲	瞽	詁	栢	楨	筊
C	蝌	霍	盥	鶴	鉸	佝	偻	媾	媪	晷	甌	覲	颶	掬	跪	香
D	滢	岌	笈	霸	緒	恠	糯	喃	曩	迺	佻	寧	孥	鬧	靸	怛
E	闌	韃	党	蟠	確	菟	咄	垌	肚	蚪	膝	灯	鐙	犖	欄	埽
F	榔	涼	魎	癘	膺	蠡	鑪	櫟	蠶	哈	洽	苓	蛉	潦	鱸	

Row 9B	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A		輻	漉	祿	轆	籟	蕾	誅	寥	儂	體	窿	凜	廩	提	漓
B	离	螭	魍	痲	豈	媽	墓	謾	鎔	鬻	魍	昧	苒	脉	俛	耄
C	銖	濛	蒙	瀾	糜	閩	婆	樽	膊	胖	髡	魴	潘	翻	辟	汴
D	辮	遍	駢	駢	迸	鴉	黼	蝠	丰	葑	仆	罨	苻	蜉	頰	鮒
E	吟	蕘	市	髯	馘	妣	沙	泚	脾	篲	腓	轡	髀	擯	蕘	贊
F	顰	鬢	似	躄	楂	槎	筍	槩	鑠	敵	錘	雲	殤	鯢	噬	

Row 9C	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A		婿	齟	射	單	尠	愴	洗	燹	霰	俛	契	挈	艘	霄	蟀
B	瑣	叟	宿	晬	洩	崇	睢	鶉	虱	啻	寺	總	矧	瑟	萼	贗
C	鴈	戛	遏	岳	諳	黯	鞅	欸	皚	礙	覿	阨	射	籥	籥	胖
D	敌	恚	射	羨	臙	蠕	讌	塩	饜	嬰	睨	翳	葦	遨	媪	蕓
E	鯢	蕘	春	蛹	吁	熨	蜿	嶂	衛	囿	帷	蕘	黝	鸞	穩	琅
F	齷	挹	楫	頤	仞	皆	棗	籽	醬	赭	柞	潛	蚕	嶂	瘡	

Row 9D	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A		装	纒	柢	蛆	豬	糴	轉	巔	磚	翦	窃	接	梃	哲	琨
B	踉	吊	噪	阜	竈	笮	糶	條	鯛	樅	蹠	蒺	肘	隼	則	証
C	恚	浪	慚	荆	悵	警	簣	蜴	躑	濺	簣	蜻	砌	隸	峭	綃
D	鈔	鞞	數	躅	躅	忽	葱	摧	槌	毳	箒	鞞	鹿	麤	杻	舳
E	冲	橈	贅	卮	徵	鴉	忱	拆	橐	馱	幘	蝠	庖	炮	舖	鉤
F	鞞	分	莧	蹕	瘡	邯	鸚	炕	靡	慊	篋	陝	皐	翻	鶻	

Row 9E	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A		鑊	園	權	鬢	媼	悅	囂	猴	篔	忻	迄	頤	刍	𪗇	𪗈
B	𪗉	𪗊	𪗋	𪗌	𪗍	𪗎	𪗏	𪗐	𪗑	𪗒	𪗓	𪗔	𪗕	𪗖	𪗗	𪗘
C	𪗙	𪗚	𪗛	𪗜	𪗝	𪗞	𪗟	𪗠	𪗡	𪗢	𪗣	𪗤	𪗥	𪗦	𪗧	𪗨
D	𪗩	𪗪	𪗫	𪗬	𪗭	𪗮	𪗯	𪗰	𪗱	𪗲	𪗳	𪗴	𪗵	𪗶	𪗷	𪗸
E	𪗹	𪗺	𪗻	𪗼	𪗽	𪗾	𪗿	𪘀	𪘁	𪘂	𪘃	𪘄	𪘅	𪘆	𪘇	𪘈
F	𪘉	𪘊	𪘋	𪘌	𪘍	𪘎	𪘏	𪘐	𪘑	𪘒	𪘓	𪘔	𪘕	𪘖	𪘗	𪘘

Row 9F	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A		𪗉	𪗊	𪗋	𪗌	𪗍	𪗎	𪗏	𪗐	𪗑	𪗒	𪗓	𪗔	𪗕	𪗖	𪗗
B	𪗘	𪗙	𪗚	𪗛	𪗜	𪗝	𪗞	𪗟	𪗠	𪗡	𪗢	𪗣	𪗤	𪗥	𪗦	𪗧
C	𪗨	𪗩	𪗪	𪗫	𪗬	𪗭	𪗮	𪗯	𪗰	𪗱	𪗲	𪗳	𪗴	𪗵	𪗶	𪗷
D	𪗸	𪗹	𪗺	𪗻	𪗼	𪗽	𪗾	𪗿	𪘀	𪘁	𪘂	𪘃	𪘄	𪘅	𪘆	𪘇
E	𪘈	𪘉	𪘊	𪘋	𪘌	𪘍	𪘎	𪘏	𪘐	𪘑	𪘒	𪘓	𪘔	𪘕	𪘖	𪘗
F	𪘘	𪘙	𪘚	𪘛	𪘜	𪘝	𪘞	𪘟	𪘠	𪘡	𪘢	𪘣	𪘤	𪘥	𪘦	𪘧

Row A0	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
A		𪗉	𪗊	𪗋	𪗌	𪗍	𪗎	𪗏	𪗐	𪗑	𪗒	𪗓	𪗔	𪗕	𪗖	𪗗
B	𪗘	𪗙	𪗚	𪗛	𪗜	𪗝	𪗞	𪗟	𪗠	𪗡	𪗢	𪗣	𪗤	𪗥	𪗦	𪗧
C	𪗨	𪗩	𪗪	𪗫	𪗬	𪗭	𪗮	𪗯	𪗰	𪗱	𪗲	𪗳	𪗴	𪗵	𪗶	𪗷
D	𪗸	𪗹	𪗺	𪗻	𪗼	𪗽	𪗾	𪗿	𪘀	𪘁	𪘂	𪘃	𪘄	𪘅	𪘆	𪘇
E	𪘈	𪘉	𪘊	𪘋	𪘌	𪘍	𪘎	𪘏	𪘐	𪘑	𪘒	𪘓	𪘔	𪘕	𪘖	𪘗
F	𪘘	𪘙	𪘚	𪘛	𪘜	𪘝	𪘞	𪘟	𪘠	𪘡	𪘢	𪘣	𪘤	𪘥	𪘦	𪘧

DBCS-PC 0x9AA6 has a glyph identical to most implementations of KS X 1001:1992 01-33 (specifically, ≠) because the actual KS X 1001:1992 standard uses the glyph ≯ for 01-33. This symbol's meaning is "not equal."

Unified Hangul Code

Microsoft introduced an expanded Korean character set, called Unified Hangul Code (UHC), beginning with Microsoft Windows 95K. Unified Hangul Code is also known as Microsoft Code Page 949.* UHC's encoding is identical to EUC-KR encoding in terms of how KS X 1001:1992 characters are encoded, but includes the additional 8,822 hangul necessary to complete the set of 11,172 hangul in the Johab set.

* Unified Hangul Code was originally called Extended Wansung.