

Adobe® FrameMaker 6.0



Chinese and Korean Features



Contents

Chinese and Korean Features	Character sets and encoding methods	3
	Inline input	4
	Typesetting rules	4
	Combined Asian and Western fonts	4
	Text import and export	5
	Date and time	5
	Autonumbering	5
	Index sorting	5
	Exporting Chinese, Korean, or Japanese documents to HTML or XML	7
	SGML	8



Chinese and Korean Features

Simplified Chinese, Traditional Chinese, and Korean features are now supported in FrameMaker® products. You can display, input, print, search and replace, import, and export Chinese and Korean text. To use these new Chinese and Korean features, install the English version of your FrameMaker product on Chinese or Korean operating systems. Version 6.0 does not provide translated interfaces and translated documentation for the Chinese and Korean languages.

The following sections describe the Chinese and Korean features in detail. For information on Asian-language filters, see the *Using Filters* online manual. For information on using Japanese characters, see the *FrameMaker User Guide*.

Character sets and encoding methods

Simplified Chinese

FrameMaker supports the GBK character set, which is a superset of the GB2312-80 character set. Only the Simplified Chinese version of Windows 95, 98, 2000 and NT supports GBK. UNIX and Macintosh support the standard GB2312-80 character set. The characters in the GBK character set are encoded in the 0x8140 and 0xFEFE code range on Windows. The characters on Macintosh and UNIX are encoded within 0xA1A1 and 0xFEFE. Because the characters between 0xA1A1 and 0xFEFE are identical on all platforms, documents are cross-platform compatible. However, if you use extended characters in GBK, they will be displayed as meaningless characters on Macintosh and UNIX.

Traditional Chinese

FrameMaker supports the Big5 character set. These characters are encoded within 0xA140 and 0xFEFE on Windows and Macintosh. Chinese UNIX supports a larger character set (CNS11643-1992) in seven code planes. The first two code planes include the same characters as in the Big5 character set, although the code mapping is different. FrameMaker supports only the first two code planes on UNIX. If you enter characters in code plane 3 and above, they will appear as spaces in FrameMaker. The traditional Chinese version of the UNIX operating system uses EUC-CNS encoding. FrameMaker provides code conversion between Big5 and EUC-CNS.

Korean

FrameMaker supports the KSC 5601-1992 character set. These characters are encoded in the 0xA1A1 and 0xFEFE code range on Windows, Macintosh, and UNIX. This encoding method is known as Wansung encoding. Windows 95 and NT 4.0 have additional Hangul characters (Windows codepage 1361), which are called Johab characters. Because Johab characters are not commonly used and are not standard, FrameMaker products do not support these characters. If you enter a Johab character, it may become two separate, meaningless single-byte characters.

Inline input

Chinese and Korean character sets contain thousands of characters, many more than the keys on a standard keyboard. To enter these thousands of characters from a standard keyboard, you use a front end processor (FEP), which is also known as an Input Method Editor (IME).

Roman letters are used to make up the phonetic pronunciation or make up the stroke or pictorial of Chinese characters. Korean characters are typically composed inline by combining basic Hangul building blocks called Jamo.

FrameMaker supports inline (on-the-spot) input methods for all text. This means you can type Chinese or Korean text directly into documents or dialog boxes.

In the Equation Editor, the inline input method is not available. You can enter Asian text using the bottom-line or root window input methods. For details on inserting a text string in an equation, see Chapter 20, “Equations” in the *FrameMaker User Guide*.

Typesetting rules

FrameMaker defines the typesetting rules for Chinese and Korean text in the *Kumihan* tables in the MIF file (see the *MIF Reference* online manual for details). The Kumihan specification defines the line-breaking rule and inter-character spacing rules for Japanese characters. FrameMaker has implemented similar rules for Chinese and Korean documents.

Asian hyphenation: A rule to prohibit certain characters from beginning a new line or ending a line is defined in the Kumihan table. You can customize the Kumihan table by modifying the table in MIF. See the Kumihan tables section in the *MIF Reference* online manual.

Spacing settings for Asian punctuation characters: The spacing settings for Asian punctuation characters, brackets, and so forth are defined in the Asian Punctuation text box in the Asian properties of the paragraph designer.

Western/Asian word spacing and Asian character spacing: These settings can be defined in the Asian properties of the paragraph designer. For details on spacing and punctuation settings, see the *FrameMaker User Guide*.

Combined Asian and Western fonts

Combined fonts assign two component fonts to one combined font name. This is done to handle both an Asian font and a Western font as though they are in one font family. In a combined font, the Asian font is the base font and the Roman font is the Western font. See the *FrameMaker User Guide* for details.

- If an Asian-language document with combined fonts is opened on a system that uses a different Asian language or a Western language, the Western component font is used for all text with the combined font. Text that used the Asian component font will be unreadable. If the document is then saved and reopened on a system with its original language, the Western text will appear correctly, but the information about the original Asian text will be lost.
- If you intend to move your documents across different Asian languages, do not use Asian characters for paragraph and character tag and combined font names. If you do, unexpected loss of data may occur.

- When you create a new document, two combined fonts are predefined in the new document. The names of the combined fonts are FMM yungjo and FMGothic f or Korean, FMSongTi and FMH eiTi for Simplified Chinese, and FMS ungTi and FMH eiTi for Traditional Chinese. The most commonly used Roman and Asian fonts are assigned as the component fonts for each combined font.

Text import and export

You can filter Chinese and Korean documents. For more information, see the *Using Filters* online manual.

Date and time

No specialized building blocks are provided for date and time variables. Day name and AM/PM are displayed in Chinese and Korean. Units for year, month, day, hour, and minute are translated in the standard template. To customize units, see Chapter 7, “Variables” in the *FrameMaker User Guide*.

Autonumbering

Full-width alphabetic characters (Western alphabets using Asian full-width fonts), Arabic numbers, and Chinese numbers are supported for paragraph autonumbers, page numbers, and footnote numbers as follows:

- <full-width a> indicates full-width lowercase alphabetic characters.
- <full-width A> indicates full-width uppercase alphabetic characters.
- <full-width n> indicates full-width Arabic numerals.
- <chinese n> indicates Chinese numerals.

Index sorting

Simplified Chinese

Index sorting for Simplified Chinese is based on the pinyin method of spelling Chinese characters using Western alphabetic characters. This phonetic method is based on Mandarin. There are nearly 400 pinyin sounds. Four tonal marks can be placed over six vowels; the tonal marks can also be represented by the numbers 1, 2, 3, and 4.

Following are examples of pinyin sorting.

Sorted Chinese word	Pinyin with accent marks	Pinyin with numbers
临别	línbié	lin2bie2
临终	línzhōng	lin2zhong1
另外	lìngwài	ling4wai4

FrameMaker products assign the most commonly used pinyin sound to each Chinese character. If the assignment is incorrect, you can specify the correct pinyin by enclosing it in brackets after the Chinese character in the index marker text. You must use numbers to represent the tonal marks, as in the following example:

行列 [hang2lie4]

A new index keyword, <\$pinyin>, is added to the SortOrderIX paragraph on the IX reference page. You cannot redefine the sort order.

Group titles for index entries are defined in the GroupTitlesIX paragraph on the IX reference page. In Simplified Chinese documents, the default group titles are the same as in Western-language documents: Symbols, Numerics, and the letters A through Z.

Traditional Chinese

The stroke-radical sort method is used for the index sorting for Traditional Chinese in FrameMaker. With this method, the number of strokes is used as the primary criterion and the type of radical is used as the secondary sort key. The sort order of radicals is based on the Kangxi Radical chart.

- A new index keyword, <\$stroke>, is added to the SortOrderIX paragraph on the IX reference page. You cannot redefine the sort order.
- Group titles for index entries are defined in the GroupTitlesIX paragraph on the IX reference page. In Traditional Chinese documents, the default group titles are as follows:

```
Symbols[ \ ]; Numerics[ 0 ]; A; B; C; D; E; F; G; H; I; J; K; L; M; N; O; P; Q; R; S; T; U; V; W; X;  
Y; Z; 一畫; 二畫; 三畫; 四畫; 五畫; 六畫; 七畫; 八畫; 九畫; 十畫; 十一畫; 十二畫; 十三畫; 十四畫; 十五畫; 十六畫; 十七畫;  
十八畫; 十九畫; 二十畫; 二十一畫; 二十二畫; 二十三畫; 二十四畫; 二十五畫以上
```

Korean

The Korean language uses Hangul (phonetic) characters and Hanja (Chinese) characters. A Hangul character is a single syllable created by combining an initial consonant, a medial vowel, and sometimes a final consonant; sorting is based on these elements in the order they occur.

The sort order of Hangul consonants is as follows:

ㄱ ㅋ ㆁ ㄷ ㅌ ㄴ ㄹ ㄷㅌ ㅌㄷ ㄷㄴ ㄴㄷ ㄷㄹ ㄴㄹ ㄷㄹ ㄴㄹ ㄹㄷ ㄹㄴ ㄹㄹ ㅁ ㄷㅁ ㅁㄷ ㄴㅁ ㅁㄴ ㄷㅁ ㅁㅌ ㄴㅌ ㄹㅁ ㅁㄹ ㄷㅁ ㅁㄴ ㄴㅁ

The sort order of Hangul vowels is as follows:

ㅣ ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅟ ㅝ ㅞ ㅟ ㅠ ㅢ ㅣ ㅤ ㅥ ㅦ ㅧ ㅨ ㅩ ㅪ ㅫ ㅬ ㅭ ㅮ ㅯ ㅰ ㅱ ㅲ ㅳ ㅴ ㅵ ㅶ ㅷ ㅸ ㅹ ㅺ ㅻ ㅼ ㅽ ㅾ ㅿ ㅿ ㅿ

The sort order of Hanja characters is determined by pronunciation. To specify the sorting of Hanja characters, add Hangul characters enclosed in brackets after the Hanja characters in the index marker text, as in the following example:

韓國 [한국]

Group titles for index entries are defined in the GroupTitlesIX paragraph on the IX reference page. In Korean documents, the default group titles are as follows:

```
Symbols[ \ ] ; Numerics[ 0 ] ; A ; B ; C ; D ; E ; F ; G ; H ; I ; J ; K ; L ; M ; N ; O ; P ; Q ; R ; S ; T ; U ; V ; W ; X ; Y ; Z ; ㄱ ; ㄴ ; ㄷ ; ㄹ ; ㅁ ; ㅂ ; ㅅ ; ㅇ ; ㅈ ; ㅊ ; ㅋ ; ㆁ ; ㆅ ; ㆆ
```

A new index keyword, <\$hangul>, is automatically added to the SortOrderIX paragraph on the IX reference page when the index is created. You can specify your own sort order for Hangul characters by replacing <\$hangul> with actual characters in the order you want them to sort.

Asian languages

If you modify the group title in an index file that was created on a different platform (for example, open an IX file generated in Windows and edit the group title in UNIX) and generate the index file again, the modified group title will not be reflected in the new index file. To resolve this problem, do not close the index file while generating the index, or turn off the Remember Missing Font Names option in the Preference dialog before opening the document from which you want to generate indexes.

Accented European characters and symbols cannot be used with Chinese or Korean characters because the same code values are used to represent multibyte characters.

For details on index generation, see “Chapter 13: Tables of Contents and Indexes” in the *FrameMaker User Guide*.

Exporting Chinese, Korean, or Japanese documents to HTML or XML

The character encoding for exporting HTML or XML is determined by the Export Encoding and CSS Export Encoding settings in the Options table on the HTML or XML reference page. For best results, both of these options should have the same settings.

The default encoding settings are:

- Japanese: Shift_JIS
- Korean: EUC-KR
- Simplified Chinese: EUC-CN
- Traditional Chinese: Big5
- Other languages: ISO-8859-1

In addition to the default encodings, you can specify ISO-8859-1 (for Western European-language systems) or UTF-8 (for browsers that support Unicode).

The Mapping table also allows encoding names used in FrameMaker (EUC-CNS for Traditional Chinese, GB for Simplified Chinese), FrameMaker+SGML (JIS8_EUC for Japanese, GB8_EUC for Simplified Chinese, KSC8_EUC for Korean), and MIME character set attributes (EUC-JP for Japanese, EUC-TW for Traditional Chinese).

The Japanese standard templates create a cascading stylesheet with Japanese tag names that are not recognized by some browsers. To resolve this problem, save the file with the default Shift_JIS encoding instead of UTF-8, use an English paragraph or character tag name, and use an English font name for the Japanese font if possible. If you modify Chinese and Korean templates and use Chinese and Korean paragraph and character tags, the same problem will occur. To resolve this problem, follow the same steps as in Japanese, using the appropriate default encoding.

Use the default encodings on Chinese, Japanese, and Korean operating systems to avoid JavaScript error messages when an HTML file is opened in Navigator™ 4.x.

In general, filenames with multibyte characters are not recommended for HTML and XML files.

SGML

You can import and export Chinese and Korean SGML files with FrameMaker+SGML. Note the following:

- Multibyte characters are not allowed in attribute and element names.
- Multibyte characters in variable names may not correctly convert into an entity in SGML.
- (UNIX) You can view Asian characters in the Structure View, but you must use the Document Window to edit.

MIF statement and keywords

To specify the encoding that was used when writing MIF statements on Chinese and Korean systems, a `MIFEncoding` statement is added at the top of Chinese and Korean MIF files:

- On Chinese systems, the statement indicates either traditional or simplified encodings:

```
<MIFEncoding ` 中文 '> #originally written as Traditional Chinese (Big5/EUC-TW)
```

or

```
<MIFEncoding ` 中文 '> #originally written as Simplified Chinese
```

The two Chinese characters in the statement mean “Chinese.” The hexadecimal representation for these two characters is A4A4 A4E5 in Big5, C4E3 C5F3 in EUC-TW, and D6D0 CEC4 in EUC-CN.

- On Korean systems, the statement is:

```
<MIFEncoding ` 한국어 '> #originally written as Korean
```

The three Korean characters in the statement mean “Korean.” Their hexadecimal representation is C7D1 B1B9 BEEE.

On Chinese and Korean systems, four MIF keywords are used to specify the numbering style of autonumber formats:

- `<FWLCAAlpha>` indicates full-width lowercase (Western) alphabetical characters.

- <FWUAlpha> indicates full-width uppercase alphabetical characters.
- <FWArabic> indicates full-width Arabic numerals.
- <ChineseNumeric> indicates Chinese numerals.

See “Autonumbering” on page 5 for the corresponding building blocks.

Other notes

PDF export for Chinese and Korean text is not supported with this version of FrameMaker.

Rubi is a Japanese system for representing the pronunciation of words as a string of phonetic characters directly above the word in question. This feature is available, although it is rarely used in Chinese and Korean text composition.

You cannot use multiple Asian languages in a single document.

When you create a new Chinese or Korean document or open a Chinese or Korean text file using the File > Open command, font and language settings are properly defined for Chinese or Korean. However, documents generated from this file (for example, Document Compare, index, TOC, SGML Error Log Report) will use the English settings, and may not display Asian characters correctly. To resolve this problem, change the English fonts to Asian fonts if the generated document is editable.

If the source document uses combined fonts, do not insert cross-references with text that includes Smart Quotes. The quotation marks become meaningless characters, and FrameMaker may crash when you use Find/Change to search for a cross-reference with Smart Quotes in a combined font.