



What is an IME (Input Method Editor) and how do I use it ?

Published: July 15, 2003

by Russ Rolfe

Latin based languages (English, German, French, Spanish, etc.) are represented by the combination of a limited set of characters. Because this set is relatively small, most languages have a one-to-one correspondence of a single character in the set to a given key on a keyboard. When it comes to East Asian (EA) languages (Chinese, Japanese, Korean) the number of characters to represent the language can be in the tens of thousands, which makes the using a one-to-one character to keyboard key model next to impossible. To allow for users to input EA characters, several Input Methods (IM) have been devised to create Input Method Editors (IME). This document's purpose is to catalog the many IMEs that exist in Windows XP and give a brief description of each.

[Windows XP](#)

[Japanese IMEs](#)

[Korean IMEs](#)

[Simplified Chinese IMEs](#)

[Traditional Chinese IMEs](#)

[Summary](#)

Windows XP

Windows has for years used Input Method Editors to allow users to enter the thousands of characters needed to write Chinese, Japanese and Korean. An IME is a program that allows computer users to enter complex characters and symbols, such as Japanese characters, using a standard keyboard. The user composes each character in one of several ways: by **radical**, by stroke count, by phonetic representation, or by typing in the character's numeric encoding index. Windows XP ships with standard IMEs that are based on the most popular input methods used in each target market. There are four basic target markets: Japanese, Korean and Chinese (which is subdivided into **Traditional** and **Simplified**).

Glossary

Radical: A radical is a group of strokes in a Chinese character that is treated as a unit for the purposes of sorting, indexing, and classification. A character can contain more than one element that is recognized as a radical, but each character contains only one element, called the "main radical," that is used as the indexing radical. The main radical often gives a hint as to the general meaning of the character, and other radicals in the character might indicate how the character is pronounced. For instance the radical “言” (which represents a word) is found in the Japanese character “go” (a language): 語.

Traditional Chinese is the set of Chinese characters, used in such countries/regions as Hong Kong SAR, Macau SAR, and Taiwan, that is consistent with the original form of Chinese ideographic characters that are several thousand years old.

Simplified Chinese is used in the People's Republic of China and Singapore. It consists of several thousand ideographic characters that are simplified versions of traditional Chinese characters.

Japanese IMEs

In order to begin entering Japanese characters in an application running on Windows XP, you need to activate the IME by selecting it from the list of input languages. When you activate the IME, the floating Language bar changes to the Japanese IME toolbar as you see in Figure A-1. Table A-1 shows what happens when you enter Japanese characters into an application running on Windows XP using the most popular phonetic input device.



Figure A-1: The Japanese IME Language Bar

Action	Result
Type the letter "k." The IME conversion (or composition) window is represented by a dotted underline. The window might be displayed anywhere on the screen, but in most applications it is displayed next to the insertion point, as in the example on the right.	<u>k</u>
Now type the letter "a." The letter "k" is replaced with the hiragana syllable "ka." If you had typed the letter "i" instead of the letter "a," the hiragana syllable "ki" (き) would have appeared.	<u>か</u>
To convert the syllable "ka" into kanji, press the Spacebar.	<u>何</u>
Suppose you are looking for a different kanji representation of "ka." If the character you are seeking is not displayed, you can activate a list of alternatives from the IME by pressing the Spacebar a second time. You can scroll down through this list (known as the "candidate window") by pressing the Spacebar a third time, as shown in the example on the right. After you have highlighted the character you want, press Enter to place it in your document. The IME responds by sending the character to the active window. Then the dotted underline representing the IME composition window disappears.	<div><u>何</u> 1 何 2 柯 3 加 4 賀 5 力 全効効ナ(KA) 6 河 7 科 8 課 9 家</div>

Table A-1: Entering Japanese characters in an application running on Windows XP.

You can form a number of kanji characters before pressing Enter. The IME engine will attempt to convert your keystrokes into a "determined string" based on Japanese grammar rules. The following explains more about the two IMEs for Japanese input in Windows XP.

Microsoft IME Standard 2002:

The standard IME that comes with Windows XP uses the above mentioned phonetic process as its default method of inputting Japanese characters. This has been the tried and true method used for many years (at least 10 yrs) to input Japanese. Beside this method, the Standard IME has what is called the IME Pad that has several other ways for users to input Japanese characters. They are:

Soft Keyboard

This is an on screen keyboard that duplicates the hardware keyboard. One uses the mouse pointer to choose which key/characters to enter. The soft keyboard has several other configurations that the user can use. Configurations are:

- Alpha/Numeric (QWERTY layout). This allows the entry of Latin characters. The only conversion that happens is Latin characters changing to/from half size and full size Latin characters.



- Alpha/Numeric (ABC layout). The soft keyboard does the same as the QWERTY layout, but it displays the characters in alphabetic order



- Hiragana/Katakana (JIS layout). This layout displays the Hiragana or Katakana characters as defined by the Japanese Industrial Standard (JIS). Once these characters are entered you can then use the conversion routines to change them from the kana to the Kanji, just like the default phonetic method mentioned above.



- Hiragana/Katakana (Syllabary layout). This layout is the standard table taught to elementary children when they are first learning the kana characters. Note that the table is read top-to-bottom and from right to left. This too allows for the conversion of the kana to kanji characters just as the JIS layout soft keyboard does.

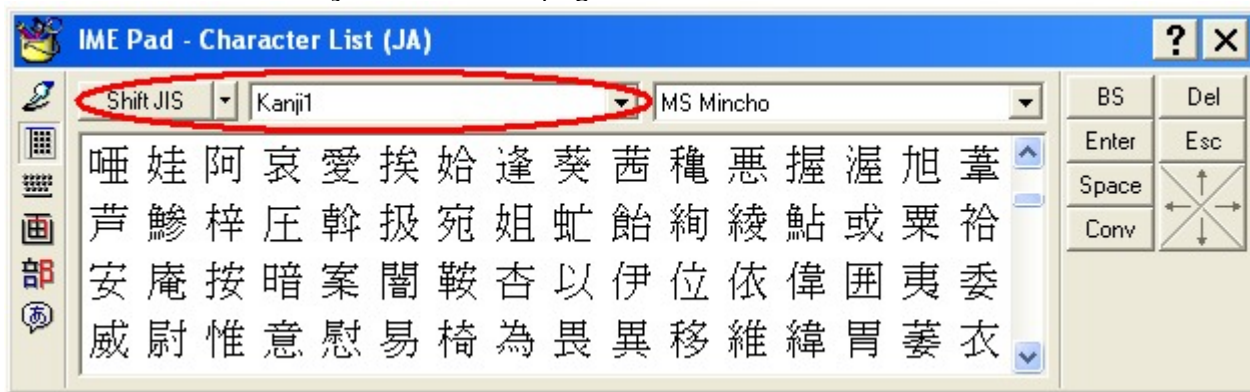


- Numeric/Date and Time. This layout gives the user access to the Kanji used for dates and time. No phonetic conversion takes place.

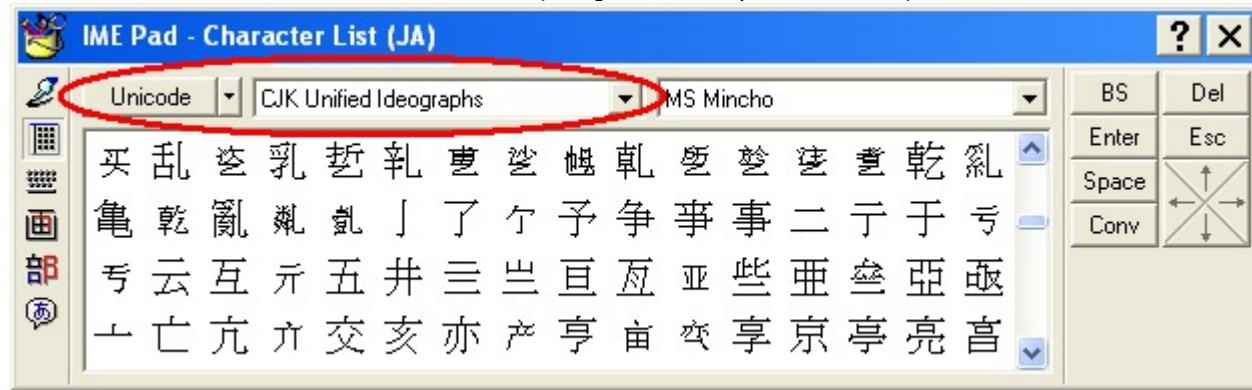


Character List

This function allows the user to choose what character to input from an interactive list. There are two different lists. The first is the Shift JIS list which displays the characters defined by the 932 code page, which is derived from the JIS standard.



The second list displays the characters defined by the Unicode standard. The extra feature that the Unicode standard list has over the Shift JIS list is it gives the user access to all of the characters (not just the Japanese ones) that Unicode defines.



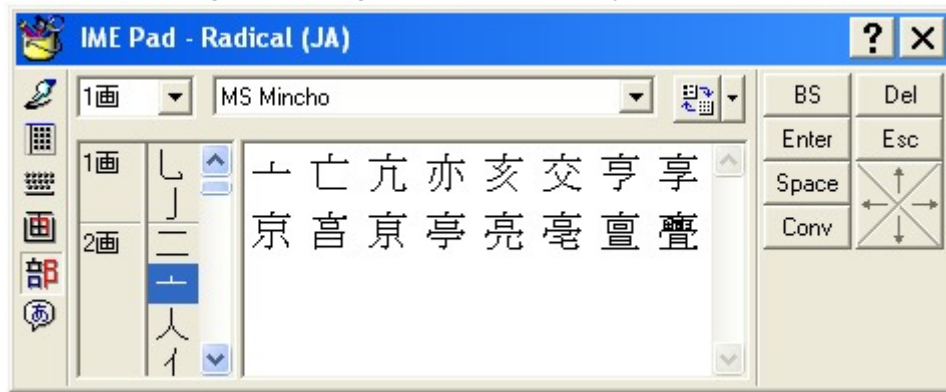
Strokes

The stroke function of the IME pad allows the user to look up a Japanese character by the number of strokes it takes to write the character. This is possible since individual Japanese characters are always written with the same number of strokes, no matter who writes them.



Radicals

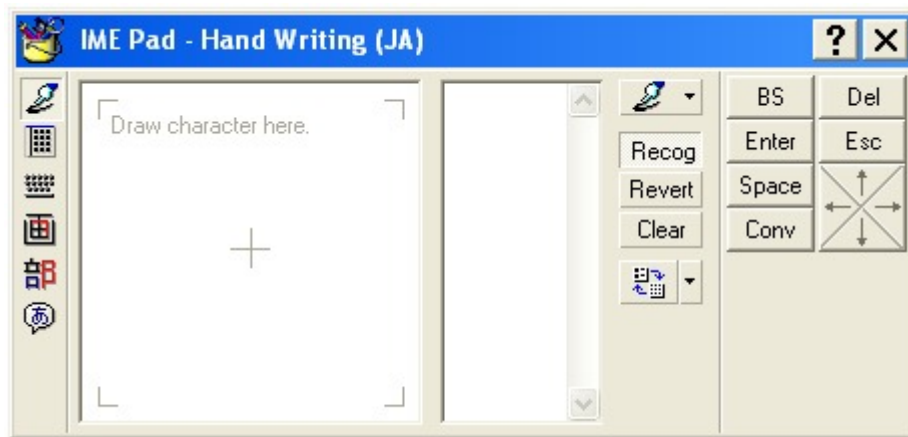
The radical function of the IME pad allows the user to look up a Japanese character in groups of main radicals (see footnote 1). Notice below how these characters are grouped together because they all have the same main radical “亠”.



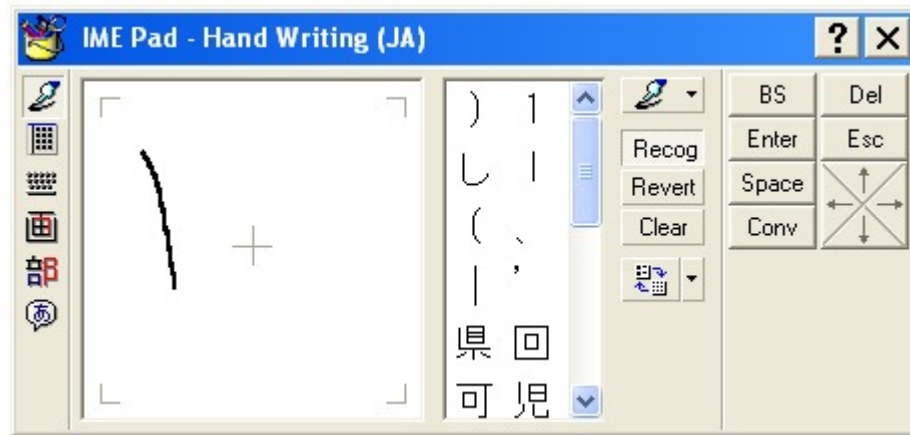
Hand Writing

The handwriting applet of the IME Pad allows one to search for and input kanji that they don't know the readings for. The IME tries to recognize the character after each stroke is written. The following four graphics show the process to enter the kanji for mouth “口”. Notice how the alternate list in the middle changes as a new stroke is added.

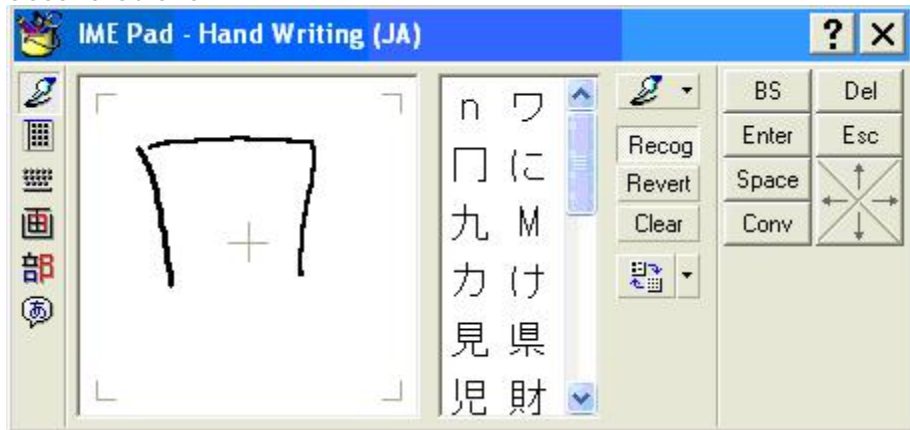
- Blank IME Pad



- First Stroke



- Second Stroke



- Third Stroke (Kanji has been recognized –character circled in red)



Microsoft Natural Input 2002

The Microsoft natural Input 2002 IME has all the same features in the Standard IME 2002, plus a feature that allows for a more natural way to edit candidates for kanji conversion. In the standard IME, the user has to recognize which editing/input mode the IME is in. Once the user converts the hiragana string to kana-kanji string, the user cannot edit each character individually. In the Natural input editing style, the user can type/edit Japanese just the same as editing Latin based language. They can delete and add glyphs one character at a time.

Korean IMEs

The Korean IME enables users to input Korean characters (Hangul) and Chinese characters (Hanja). The Microsoft Korean IME 2002 operates in two modes. It operates in the new mode with full features when used with an application that supports this new architecture (such as Microsoft Office XP), but when it is used with a legacy application, like Notepad, it automatically switches into the old Microsoft Korean IME 2000.

Hangul is entered by representing its 24 basic elements and combination of elements, all of which are called Jamos, on the standard 101 keyboard. By combining these Jamos, all 11,172 Hangul character combinations can be produced. This is done as much as in the same way mentioned before that Japanese is entered in phonetically.



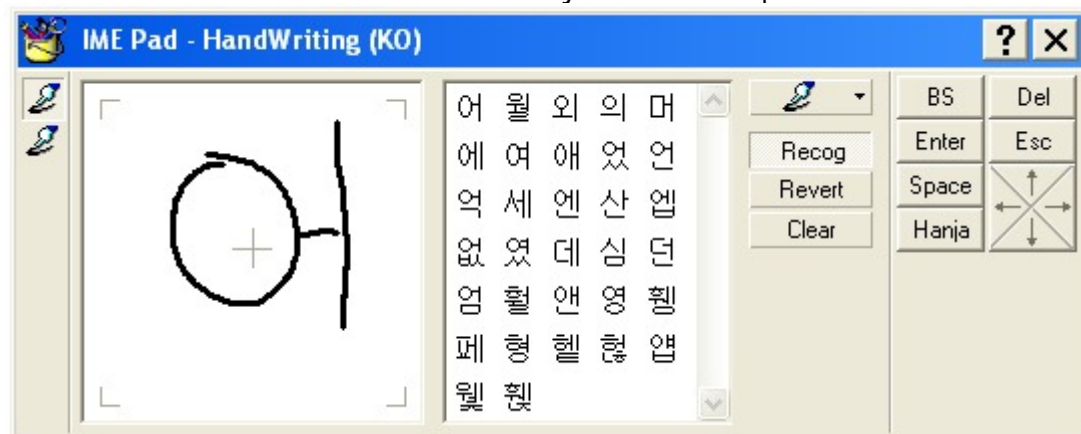
The following table (A-3) shows how the Korean character for language (언어) is entered in to an application.

Action	Result
Type the letter "ㅇ" that corresponds to English key "d."	ㅇ
Now type the letter "ㅈ" that corresponds to the English key "j." The character "ㅇ" is replaced with the combined syllable "어."	어
Type the letter "ㄴ" that corresponds to English key "s." The character combination is now replaced with "언", which finishes the first of the two hanguls needed to represent language	언
To create the second hangul, type the letter "ㅇ" again.	언ㅇ
Now type the letter "ㅈ" again followed by a space to finish the word.	언어

Once the Hangul has been formed, the user can then press a Hanja key that will allow the Hangul to be transformed into Chinese Characters called Hanja. The Korean IME has an IME Pad that allows the user to input Hangul via soft keyboard.



Plus it has a Handwriting applet that allows one to search for and input hangul, hanja and kanji (if the Japanese IME is installed). The IME tries to recognize the character after each stroke is written just like the Japanese version.



When the Japanese IME is installed, The Korean IME has access to the Japanese IME Pad's Character List, Strokes and Radical functions, which allows the IME to enter characters outside the normal set of hungul and hanja.

Simplified Chinese IMEs

Simplified Chinese is the script used in the People's Republic of China and Singapore. It consists of several thousand ideographic characters that are simplified versions of the traditional Chinese characters.

Like Japanese, Simplified Chinese characters can be entered by radical, by stroke count, by phonetic representation, or by typing in the character's numeric encoding index. The phonetic input is based on what is called **Pinyin**. It is the standard Simplified Chinese phonetic script in mainland China. It was promulgated in 1958 and has been used worldwide as an educational tool to learn the sounds of Mandarin Chinese. It has been the United Nations standard since 1977 and an ISO standard since 1982. The system uses the basic Roman alphabet to represent the Mandarin sounds. There are several input methods based on modifications to Pinyin. One notable method is called **Double-Pinyin**, where the amount of keystrokes of the pinyin input method is reduced to one or two Latin letters.

Unlike Japanese and Korean IMEs, Simplified Chinese IMEs' **composition window** is a combination of characters displayed in-line along with a reading window to display the keystrokes entered with the surrounding text. Figure A-2 shows the Japanese composition window in-line (the underlined characters). Figure A-3 shows the Chinese New Phonetic composition windows as a small box below the line. The composition window is the window of an IME that displays text typed by the user.

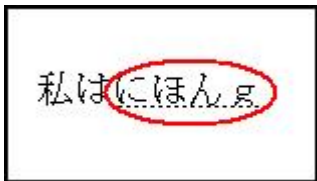



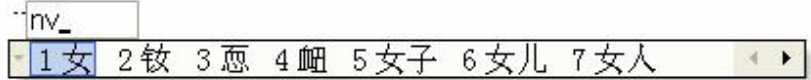



Figure A-3: Japanese IME in-line composition window.



Figure A-4: Simplified Chinese MSPY IME windows.

There are 8 IMEs in Windows XP to support Simplified Chinese. They are:

MSPY: MSPY is a pinyin-based IME that supports both pinyin input and double-pinyin input. The key advantage of MSPY is that MSPY is powered by a statistics-based language model (LM) plus linguistic rules which gives it a higher pinyin conversion accuracy than any other pinyin-based IMEs currently in China market. Besides pinyin input, the latest MSPY2003 for Office2003 supports Internal Code input too, such as Unicode/Surrogate input and GBx code input. To give one a better understanding on how this pinyin-based IME is used, the following are the steps needed to enter the Chinese characters for daughter (女儿), which in pinyin representation is "nver".

Action	
	Result
Type the letter "n". The IME conversion (or composition) window is represented by a dotted underline and the reading window is represented by the grey box. The window might be displayed anywhere on the screen, but in most applications it is displayed next to the insertion point, as in the example on the right.	
	
Now type the letter "v." The candidate window appears with several choices.	
	
Now type the letter "e". The first character appears in the composition window and a new candidate window appears with several other choices.	
	
Now type the letter "r". Another list of characters appears in the candidate windows. Notice that 1 st , 2 nd and 3 rd candidates have a set of glyphs with the first character grayed out. This is to indicate that these are possible combined combinations forming one word. You can scroll down through candidate window by clicking on the arrows at the right of the list.	
	
To enter your choice, either press the space bar to accept the 1 st choice as the default, or press the corresponding number to character/s you want. In this case it would be "1".(Note that in step 2, choice 6 is what we were looking for, so we could have type in the number "6" at that time.) The IME enters your choice in to the composite window and closes the reading and candidate windows. By pressing the space bar again, the IME responds by sending the character to the active window. Then the dotted underline representing the IME composition window disappears.	
	

Besides entering characters via pinyin from the computers keyboard, MSPY has 12 other soft keyboards one can use. They are:



PC Keyboard



Greek



Russian

`	1	2	3	4	5	6	7	8	9	0	-	=	←
	ウ	ク	シ	ス	セ	ソ	タ	チ	ツ	テ	ト	ニ	
Tab	Q	W	E	R	T	Y	U	I	O	P	[]	\
	タ	カ	キ	ク	ケ	コ	サ	シ	ス	セ	ソ	タ	
Caps	A	S	D	F	G	H	J	K	L	;	'	↵	Enter
	ア	イ	ウ	エ	オ	カ	キ	ク	ケ	コ	サ	シ	
⇧ Shift	Z	X	C	V	B	N	M	,	.	/			
	ツ	ク	セ	ソ	タ	チ	ツ	テ	ト	ニ			
Esc												Ins	Del

Phonetic

`	1	2	3	4	5	6	7	8	9	0	-	=	←
	ā	ō	ō	ō	ō	ō	ō	ō	ō	ō	ē		
Tab	Q	W	E	R	T	Y	U	I	O	P	[]	\
	ā	ō	ō	ō	ō	ō	ō	ō	ō	ō	ē		
Caps	A	S	D	F	G	H	J	K	L	;	'	↵	Enter
	ā	ō	ō	ō	ō	ō	ō	ō	ō	ō	ē		
⇧ Shift	Z	X	C	V	B	N	M	,	.	/			
	ū	ū	ū	ū	ū	ū	ū	ū	ū	ū	ū		
Esc												Ins	Del

Pinyin letter

`	あ	い	う	え	お	か	き	く	け	こ	-	=	←
	あ	い	う	え	お	か	き	く	け	こ	ん		
Tab	Q	さ	じ	ず	ぜ	そ	た	ち	つ	っ	で	[ど]	\
	さ	じ	ず	ぜ	そ	た	ち	つ	っ	で	と	ぬ	
Caps	A	ば	S	D	F	G	H	J	K	L	;	'	↵
	ば	に	ぬ	ね	の	は	ひ	ふ	へ	ほ	ろ		
⇧ Shift	Z	ら	X	り	C	る	V	れ	B	ろ	N	M	
	ま	み	む	め	も	や	ゆ	よ	わ	を			
Esc												Ins	Del

Hiragana

`	1	2	3	4	5	6	7	8	9	0	-	=	←
ア	イ	ウ	エ	オ	カ	キ	ク	ケ	コ				
Tab	Q	W	E	R	T	Y	U	I	O	P	[]	↵
サ	シ	ス	セ	ソ	タ	チ	ツ	テ	ト	キ			
Caps	A	S	D	F	G	H	J	K	L	;	'		↵
ナ	ニ	ヌ	ネ	ノ	ハ	ヒ	フ	ヘ	ホ	エ			
⇧ Shift	Z	X	C	V	B	N	M	,	.	/			
マ	ミ	ム	メ	モ	ヤ	ユ	ヨ	ワ	ヲ				
Esc											Ins	Del	

Katakana

`	1	2	3	4	5	6	7	8	9	0	-	=	←
。	、	、	、	、	、	、	、	、	、	、	、	、	
Tab	Q	W	E	R	T	Y	U	I	O	P	[]	↵
、	、	、	、	、	、	、	、	、	、	、	、	、	
Caps	A	S	D	F	G	H	J	K	L	;	'		↵
()	<	>	《	》	「	」	『	』	.			
⇧ Shift	Z	X	C	V	B	N	M	,	.	/			
【	】	()	[]	{	}						
Esc											Ins	Del	

Punctuation

`	1	2	3	4	5	6	7	8	9	0	-	=	←
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII		
Tab	Q	W	E	R	T	Y	U	I	O	P	[]	↵
1.	2.	3.	4.	5.	6.	7.	8.	9.	10.				
Caps	A	S	D	F	G	H	J	K	L	;	'		↵
(-)	(-)	(=)	(四)	(五)	(六)	(七)	(八)	(九)	(十)				
⇧ Shift	Z	X	C	V	B	N	M	,	.	/			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)				
Esc											Ins	Del	

Math number



Math symbol



Table symbol



Special symbol

When the Japanese IME is installed, The MSPY IME has access to the Japanese IME Pad's Character List, Strokes and Radical functions, which allows the IME to enter Simplified Chinese characters and characters outside the normal set of Chinese characters.

Enhanced Unicode allows one to input characters via their Unicode encoding. Instead of the pinyin entered into the reading window, the Unicode hexadecimal encoding is used. One can also input the Unicode surrogate pair code points to get their corresponding character. Enhanced Unicode supports Unicode/GBK/QuWei besides surrogate pairs.

GBK stands for the Chinese government standard character code GB13000-1. It accommodates the Chinese characters that are in ISO 10606-1:1993 that are not found in the original Government standard GB 2312-80. GKB is what the Microsoft Code-Page 963 is based.

[What is an IME \(Input Method Editor\) and how do I use it ?](#)

QuWei (translation "regional position") inputs Chinese characters using the Row-Cell numbers in GB 2310-82.

NeiMa is also an IME that allows input via an encoding schema. It is a table-driven Windows IME. It currently supports Unicode/GBK/QuWei, but it does not support Unicode surrogate pairs.

QuanPin is a pinyin-based IME that supports only full pinyin (no double-pinyin support). It is a table-driven IME. One inputs pinyin word by word to get the corresponding Chinese characters. It has a low pinyin conversion accuracy.

ShuangPin is a pinyin-based IME that supports only double-pinyin (no full pinyin support). It is a table-driven IME. One inputs double-pinyin word by word to get the corresponding Chinese characters. It has a low pinyin conversion accuracy too.

WuBi86 is a stroke-based IME used mainly by professional typist in mainland China. The best designed and perhaps the fastest input method of this kind is Wubi, devised by Wang Yongmin from mainland China. With this input method, common characters can be input with just 2 key strokes and the maximum number of key strokes for any character is four. What is really amazing about this input method is that there is almost no need to choose from a candidate list because most of the key combinations are unique. Furthermore, the mechanism extends into multiple character words and phrases and assigns the maximum of 4 key strokes to them. A well-trained and experienced user can type over 100 characters per minute, faster than one can speak the language !.

WuBi98 - WuBi98 is an updated version based on WuBi86. WuBi 86 is an old version that was released years ago. The Chinese government has a new standard of Chinese strokes, the major update of WuBi98 is based on the new stroke standard. The name **Wubi**, literally means "five strokes", but it has nothing to do with stroke number. It simply divides the keyboard into five major regions, hence the name. However, structure-based input methods are usually difficult to learn and easy to forget. In addition, you have to know how to write a character in order to type it. By contrast, transliteration input methods like pinyin are more intuitive and easy to grasp.

ZhengMa - ZhengMa is a stroke-based IME. It's a table-driven IME of Windows. The difference between ZhengMa and WuBi is the way it defines the stroke coding.

Traditional Chinese IMEs

Traditional Chinese is the set of Chinese characters, used in such places as Hong Kong SAR, Macau SAR, and Taiwan. They are consistent with the original forms of Chinese ideographic characters that are several thousand years old. There have been IMEs for Traditional Chinese even before Windows 3.1. Traditional Chinese IMEs can be broken up into two categories:

- **Legacy IMEs** have shipped since Windows 3.1 or even earlier. They are also called character based IMEs. Their basic operation is where the user types in a sequence of keys and manually chooses the Chinese character from an **alternative list** for each character. This is known as type-and-select.
- **Intelligent IME** was invented in Windows NT 4.0 time frame. They are also named sentence-based IME in that they are designed to utilize the computational linguistics technology so that users can just type the input key sequence of the character and IME will choose the best match character automatically.

The **Alternative list** is a small box, other than the composition window, containing the different possible character choices derived from the sequence of keys entered by the user.

Like Japanese, Traditional Chinese characters can be entered in by radical, by stroke count, by phonetic representation, or by typing in the character's numeric code-page index. The phonetic input is based on what is called Bopomofo. It is the standard Traditional Chinese phonetic script in Taiwan. The phonetic sounds are either represented by Latin characters or Chinese phonetic symbols. Its name comes from the first four Chinese phonetic symbols Bo (ㄅ) Po (ㄆ) Mo (ㄇ) Fo (ㄈ), which corresponds to the b, p, m, f keys on a standard 101 keyboard. Bopomofo consists of 42 symbols (21 'consonants', 16 'vowels' and 5 'tone marks').

Unlike Japanese and Korean IMEs, Traditional Chinese IMEs' composition window is a combination of characters displayed in-line along with a reading window to display the keystrokes entered. Figure A-4 shows the Japanese composition window in-line (the underlined characters). Figure A-5 shows the Chinese New Phonetic composition windows as a small box below the line.

The **composition window** is the window of an IME that displays text typed by the user.

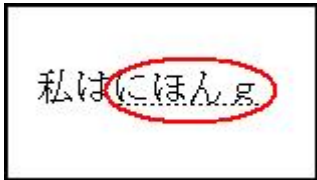


Figure A-5: Japanese IME in-line composition window.

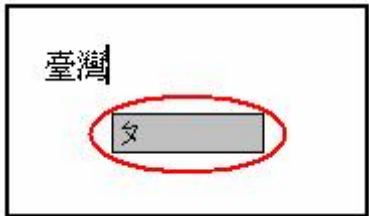


Figure A-6: Traditional Chinese IME composition and reading (grey box) windows.

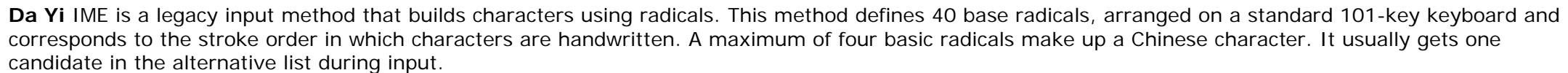
There are 10 traditional Chinese IMEs supported in Windows XP:

Alphanumeric IME is an input method for inputting full-width and half-width alphanumeric characters. It does not input Chinese characters.

Full-width character Characters whose glyph image extends across the entire character display cell. In legacy character sets, full-width characters are normally encoded in 2 or more bytes. The Japanese term for full-width characters is "zenkaku." **Half-width** character Characters whose glyph image occupies half of the character display cell. In legacy character sets, half-width characters are normally encoded in a single byte. The Japanese term for half-width characters is "hankaku."

Array IME is a legacy input method that builds characters using radicals. This method defines 10 basic sub-radicals (0-9). The columns of keys beneath each number—for example, on a US keyboard the letters QAZ beneath 1 and the letters WSX beneath 2—are used to select specific characters. Characters are composed by 2 sub-radicals plus an encoding from 00 to 99. Each Chinese character is composed by 2 radicals. It usually gets the fewest candidates in the alternative list during input.

ChangJie IME is a legacy input method that uses radicals to build Chinese characters. It usually gets one candidate in alternative list during input. It divides 24 radicals in 4 groups, the Philosophical group (represented by the Latin characters A, B, C, D, E, F, G), the Stroke group (represented by H, I, J, K, L, M, N), the Body parts group (represented by O, P, Q, R) and the Character shape group (represented by S, T, U, V, W, Y). The letter X is used to generate more complex radicals. Up to five keystrokes are required to generate a single Chinese character using the following sequence of patterns: top-down, left-right, and outside-inside. Each pattern style consists of a character heading and a character body. Building a character heading requires a maximum of two keystrokes (the first and last radicals). Building a character body requires a maximum of three keystrokes (the first, second, and last radicals). This input method adheres to the Chang Jie tables published by E-Ten Information Systems.



Phonetic IME is a legacy input method that uses phonetics to build Chinese characters. Since the 42 symbols of Bopomofo gets mapped on to the standard 101, some numbers, symbols and punctuations on the Standard English layout are not directly available when typing. The user generates Chinese characters by combining consonants, semi-vowels, and tones. It usually gets many candidates in the alternative list during input.



Unicode IME is a legacy input method that uses Unicode character encodings to input Chinese characters.

New ChangJie IME is an intelligent input method that uses radicals to build Chinese characters. It uses the ChangJie IME radicals as the initial input.

New Phonetic IME is an intelligent input method that uses phonetics (Bopomofo) to build Chinese characters. It is a superset of the Phonetic IME. It provides several popular keyboard mappings (including soft-keyboards), plus pinyin input as well. The following are its soft-keyboard layouts:

- Standard Bopomofo keyboard layout



- Eten keyboard layout



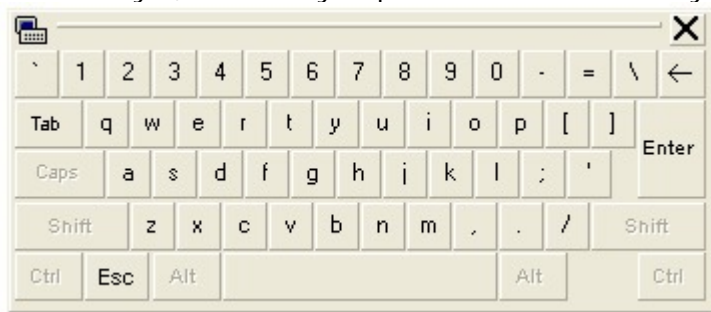
- ChingYeah keyboard layout



- IBM keyboard layout



- HanYu Pinyin, Secondary Bopomofo and Taiwan Pinyin keyboard layouts



Although the HanYu Pinyin, Secondary Bopomofo and Taiwan Pinyin keyboard layouts appear to be the same, when invoked, they each treat the entry of characters via pinyin phonetic input a little differently. They each have their own distinctive way to define the consonants, vowels and compound vowels. (See the New Phonetic IME's help for more information on these input methods.)

When the Japanese IME is installed, The New Phonetic IME has access to the Japanese IME Pad's Character List, Strokes and Radical functions, which allows the IME to enter Traditional Chinese characters and characters outside the normal set of Chinese characters.

Summary

Microsoft's Windows XP has many different ways to input East Asian characters, which allows the user to choose the one they are more comfortable with. This is just one of the areas where Microsoft excels in providing solutions for people's international needs.