SHOPPING AROUND FOR A CHINESE WORD PROCESSOR

INTRODUCTION

Rather than providing answers, this paper suggests the kind of questions to ask when selecting software. Very little if anything has been written evaluating the various packages, and I have not had the opportunity to test them myself. Therefore I prefer to indicate, in general terms, what the problems are in Chinese word processing, and what to look for.

The reason for Chinese automation having more problems than English is, of course, the writing system but there are a couple of other factors. This paper looks first at the problem, then at Chinese input methods, at word processors in general, at what is currently available, and lastly at selection criteria.

THE PROBLEM

There is a problem in manipulating and storing up to 60,000 Chinese characters as two or three byte characters in traditional, simplified or variant forms, as well as uniquely Japanese characters. As computer sizes diminish and capacity increases, this problem is decreasing.

With the size of the character set comes the difficulty of coding the characters uniquely in a form which is logical and learnable. Rather than learn codes by heart, the user should be able to apply the same rules to each character in order to key it in correctly. In fact the layout of the elements of the Chinese character is logical when looked at from the viewpoint of the history and development of the written language, but not systematic in a mathematical way.

CHINESE INPUT METHODS

Apart from numeric or hex codes which must be learned by heart, such as telegraphic code, there are two main input methods: phonetic-conversion and Chinese character structure.

The phonetic input methods include romanisation, Chinese Phonetic Alphabet (Chu yin fu hao), and, in the case of Japanese, hiragana and katakana. Phonetic-conversion involves input according to pronunciation and, in some cases optionally, a tone indicator, which the system converts to Chinese characters; where

there are ambiguities, the system offers a choice for selection. The advantages of romanisation-conversion input methods are that a standard keyboard is used, and many people have keyboard skills; also, non-native Chinese speakers are familiar with romanisation and take to it easily. On the other hand, whereas each Chinese character is unique, its romanised equivalent is not; additionally, speakers of non-standard Chinese will have difficulty with spelling.

Input by Chinese character structure involves breaking down the character into component parts, sometimes by radical, sometimes by stroke, and keying in several components or codes, which the system converts and offers the complete Chinese character for. Although the characters are usually indexed uniquely, there may be more than one character with the same component parts, and a selection may be offered to the user.

Whilst the user does not need to know standard pronunciation, he or she must learn one of many company-based systems: there is no standard radical set and this is the major reason for the high number of competing systems on the market. These methods base their approach on logical rules but unfortunately the rules do not fit every Chinese character and there may be many exceptions. Such methods involve training staff who would have to use the system regularly to maintain their typing speeds. They are not for the casual user.

A combination of these two methods is also possible; that is, the user combines pronunciation, tone and radical in order to reduce the number of ambiguities. Unfortunately it results in the disadvantages of both systems being present, i.e., the need to know the pronunciation, tone and a company-based system.

Ideally a word processor should have a choice of input methods, since the user may know the composition and not the pronunciation of some Chinese characters, but the pronunciation and not the exact composition of others.

Word processors developed in China and Taiwan tend for political or nationalistic reasons not to include the character sets of the other. They also do not usually include romanisation-conversion methods, since the Chinese do not really like the romanisation approach. A prime example of this was a McDonnell-Douglas Microdata Chinese terminal, made in Taiwan, which ANU Library tested in 1987. The good news was that it interfaced with URICA without any apparent problems and that it supported 30,000 Chinese characters and 5 input methods. The bad news was that it had no simplified characters, no other character sets such as Japanese kana, and no romanisation-conversion input method.

WORD PROCESSORS AND THEIR FEATURES

What features are standard among word processor and can we expect the same in a Chinese word processor? In fact there are generally fewer features available in Chinese mode; the prospective buyer should check that there are enough features present to make editing easy. The following are fairly standard for English-language word processors:

> Insert, delete, cut, move, paste, justify; margins, pagination, double or single line spacing, automatic reformatting, bold, italics, underline, superscript, subscript, WYSIWYG, import files.

WYSIWYG - What You See Is What You Get - means what it says: the user should be able to see on the screen how the text will appear in print. For instance, underlined characters would appear as underlined on the screen. This feature is not always present in Chinese word processors; instead the text is marked with a special character at the beginning and end of the underlined section. Other features may be similarly treated.

The capability to print both vertically and horizontally may be needed in a Chinese word processor.

WHAT IS CURRENTLY AVAILABLE IN AUSTRALIA?

The short answer is "Not much". In Taiwan and China, Chinese word processors proliferate but they are not marketed here. For reasons discussed earlier, they may not be suitable for Australian conditions and we find ourselves looking to USA for packages to handle traditional and simplified characters, more than one script or to interface with other software.

Apple has developed fonts and font packages for the Macintosh which can be used with standard word processors such as MacWrite and Microsoft Word. A Chinese package with Cantonese romanisation input method is available, and for which a Mandarin romanisation input method is currently under development. Input is usually by phonetic-conversion. There are a couple of full word processing packages for the Macintosh with additional input methods.

IBM and IBM users have developed systems rather than separate fonts, in the way of word processors and business packages with varying input methods.

Xerox has developed multi-lingual software which, from hearsay, is supposed to be extremely good but the hardware required to run it is expensive.

There are two Wang packages available in Australia; they are basically English-language database management systems which also accept Chinese characters. Wang supports the 3-corner code input method.

The two serious contenders for non-Roman script software are Apple and IBM. Apple appears to have the edge over IBM as far as cheap packages, versatility and user friendliness is concerned. On the other hand, they do not appear to have developed to the same extent input methods based on Chinese character structure.

The standard IBM PC cannot handle Chinese characters without additions such as a graphics card and high resolution monitor. In other words, while you can sit down and use Chinese and Japanese programs on most Macintosh Pluses with 2 disc drives and sufficient RAM, you cannot sit down at any IBM PC and successfully load Chinese word processing software. Even when the correct equipment configuration is available, in my experience some programs have proved difficult for the non-professional computer user to load. That was before attending beginners' and advanced DOS courses, admittedly; novices may need help from their ADP department.

IBM compatibility can also be a problem: if a package is developed on an IBM microcomputer, it may need 99% compatibility to run on other brands of microcomputers.

Software with multiple input systems tends to include a radical input method, the Chinese Phonetic Alphabet, Telegraphic Code and often a hex code. Of the many existing radical input methods, a few are fairly widely used, one of them being Tsang chieh (Cangjie) and its simplified version, Quick Tsang chieh (see Attachments).

Chinese word processing software for IBM and APPLE Macintosh:
- Brushwriter (also runs on Atari)

Apple Macintosh packages:
- Zhongwen Talk
- Feima Hard Disk SE
- MacChinese font (Cantonese romanisation input)

Word processing software for IBM or compatibles only:
- ChinaStar II products
- Duke Chinese Typist
- Hanyupinyin Word Processor
- OCLC CJK350 Word Processor
- Professional ChinaStar
- PX 2001 Pinxxiee Chinese Word Processor
- Tianma

Wang products:
- Libman (library and information management system)
- Recman (record and information management system)

For more details, readers are referred to a recently released directory of software for non-Roman scripts for descriptions of actual packages (1).

SELECTION CRITERIA

1. What is the purpose of buying this software and what should it be able to do? What printed products should it be able to produce?

2. What budgetary and hardware restrictions apply?

3. Is the hardware to be used exclusively for this software or must it be versatile? Is dedicated equipment and a non-standard keyboard acceptable?

4. Must the software be able to interface with another system or send files via electronic mail?

5. Depending on the purpose of the software:
   how many Chinese characters are needed?
   must it be able to handle traditional, simplified and variant forms and mix them in the same text?
   can it handle more than one script,
   e.g. can English and Chinese be mixed in the text?

6. Does the vendor supply support and enhancements?

7. Who will be operating the software?
   Are they likely to be native or non-native speakers?
   Do you plan for a limited number of highly-trained users or a variety of casual users who operate it for relatively small amounts of time?
   These factors affect the preferred type of input method.

8. How many input methods come with the software, and are they suited to the anticipated types of users?

9. How difficult are the input methods?

10. How much training do the operators need and who will provide it?

11. Are the manual and commands in the most suitable language for the operators?

12. Is the system tolerant of input errors; that is, can it calculate the most probable Chinese character from an erroneous input code?

13. Does the the facility exist for creating extra characters?

14. Are there sufficient editing features present?

CONCLUSION

Selecting Chinese word processing software is basically a matter of preparation and analysis of one's needs, then of comparing them with what is available. The foregoing should give some guide as to what to look for. Having chosen a word processor which most nearly fills the requirements, it is advisable to test it before deciding to purchase. Demonstrations and demonstration discs may have their place, but are no substitute for a "hands on" trial. Testing is the only way to judge a package's ease of use and suitability.

Susan MacDougall
Canberra College of Advanced Education

NOTES

(1) MacDougall, S. Non-Roman scripts: a software directory. Canberra: S. MacDougall, 1988.

BIBLIOGRAPHY

Chang, Ifay and Yu, Wellington. Survey of Chinese input methods: International Conference on Chinese Computing Tutorial Papers. Singapore: University of Singapore, Institute of Systems Science, 1986.

Chen, Chen-kuang and Gong, Reng-weng. "Evaluation of Chinese input methods." ***Computer Processing of Chinese and Oriental Languages***, vol. 1, no. 4 (November 1984): 236-247.

Lo, S.Y. A scientific model for comparing various methods of inputting Chinese characters into computer. [unpublished paper] Melbourne, 1985?

ATTACHMENT 1

TELEGRAPHIC CODE

widespread use
4-digit numeric code

from: Chang, Ifay and Yu, Wellington. Survey of Chinese input methods: International Conference on Chinese Computing Tutorial Papers. Singapore: University of Singapore, Institute of Systems Science, 1986.

ATTACHMENT 2

CHINESE PHONETIC ALPHABET

Widespread use in Taiwan and China.
37 symbols plus tone indicators occupying alphabetic, numeric and some punctuation keys on keyboard

21 initials
16 finals
tone markers

注音符號表

(Initials) 聲符表	(Finals) 韻符表	(Tones) 聲調表
ㄅ ㄆ ㄇ ㄈ	ㄧ ㄨ ㄩ	陰平或第一聲 無符號
ㄉ ㄊ ㄋ ㄌ	ㄚ ㄛ ㄜ ㄝ	陽平或第二聲 ˊ
ㄍ ㄎ ㄏ	ㄞ ㄟ ㄠ ㄡ	上聲或第三聲 ˇ
ㄐ ㄑ ㄒ	ㄢ ㄣ ㄤ ㄥ	去聲或第四聲 ˋ
ㄓ ㄔ ㄕ ㄖ	ㄦ	
ㄗ ㄘ ㄙ		

from: Chang, Ifay and Yu, Wellington. Survey of Chinese input methods: International Conference on Chinese Computing Tutorial Papers. Singapore: University of Singapore, Institute of Systems Science, 1986.

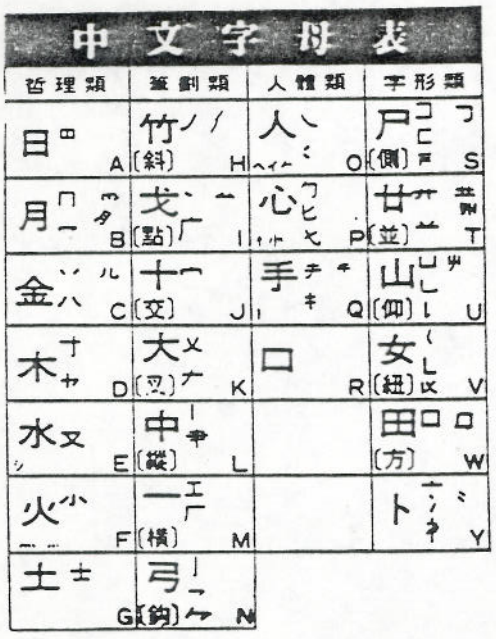
ATTACHMENT 3

TSANG CHIEH (QUICK TSANG CHIEH)

Used by at least 4 Taiwanese companies in 1986; also currently in US software such as OCLC. 24 radicals represented by letters A to W with X and Z reserved for special uses. Code can be from 1 to 5 characters long.
Quick Tsang Chieh uses the first and last code letter; the system offers a selection of possible characters.

RADICAL CHART

倉頡輸入法中文字母



BASIC PRINCIPLES

character	elements	code
明	日，月	AB
天	一，大	MK
王	一，土	MG
昌	日，日	AA
奎	大，土，土	KGG
肚	月，土	BG
針	金，十	CJ
合	人，一，口	OMR
足	口，卜，人	RYO
社	木，土	DG
炎	火，火	FF
思	田，心	WP
或	戈，口，一	IRM
箕	竹，戈，戈	HII
是	日，一，卜，人	AYO

ATTACHMENT 4

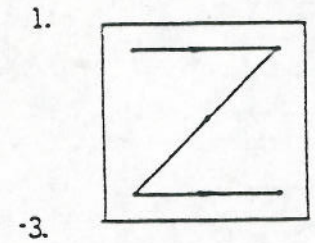
3-CORNER CODE

Used by at least 10 Taiwanese companies in 1986 including Wang Computer

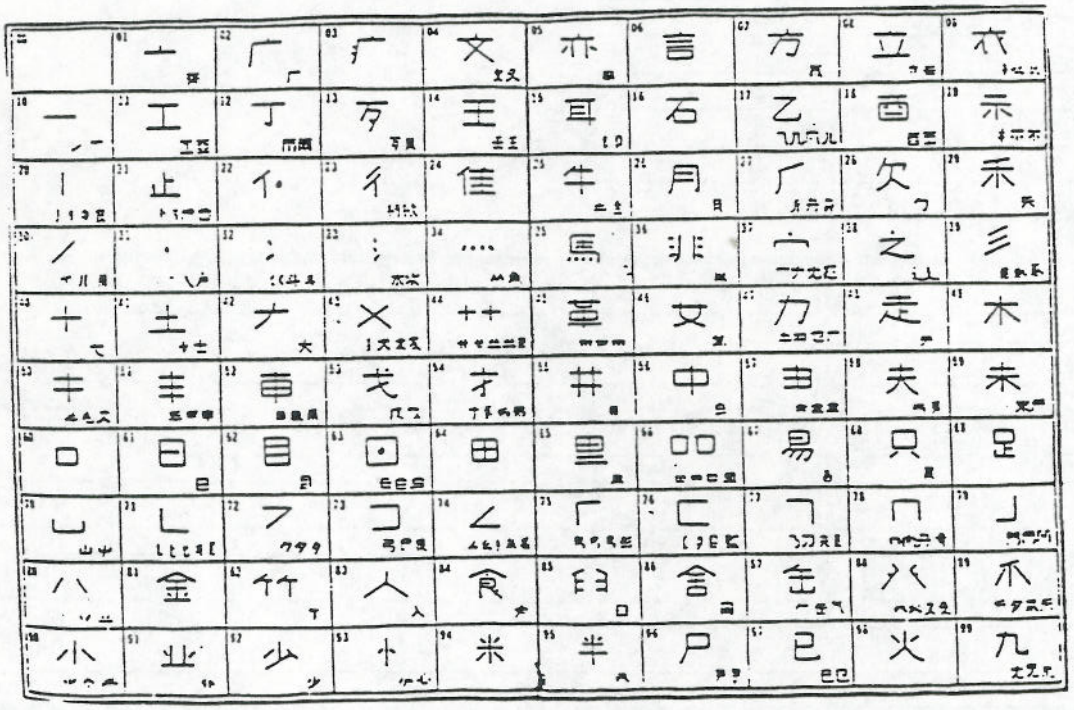
CODING RULES

取角原則的基本順序，是從左到右，由上而下，圖示如下：

1. 左上角
2. 右上角
3. 左下角
4. 右下角



RADICAL CHART
Numeric code from 1-99



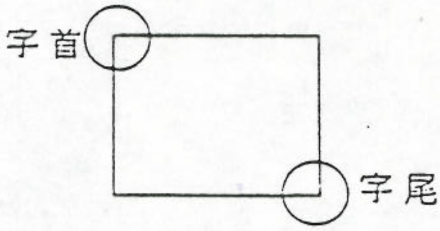
from: Chang, Ifay and Yu, Wellington. Survey of Chinese input methods: International Conference on Chinese Computing Tutorial Papers. Singapore: University of Singapore, Institute of Systems Science, 1986.

ATTACHMENT 5

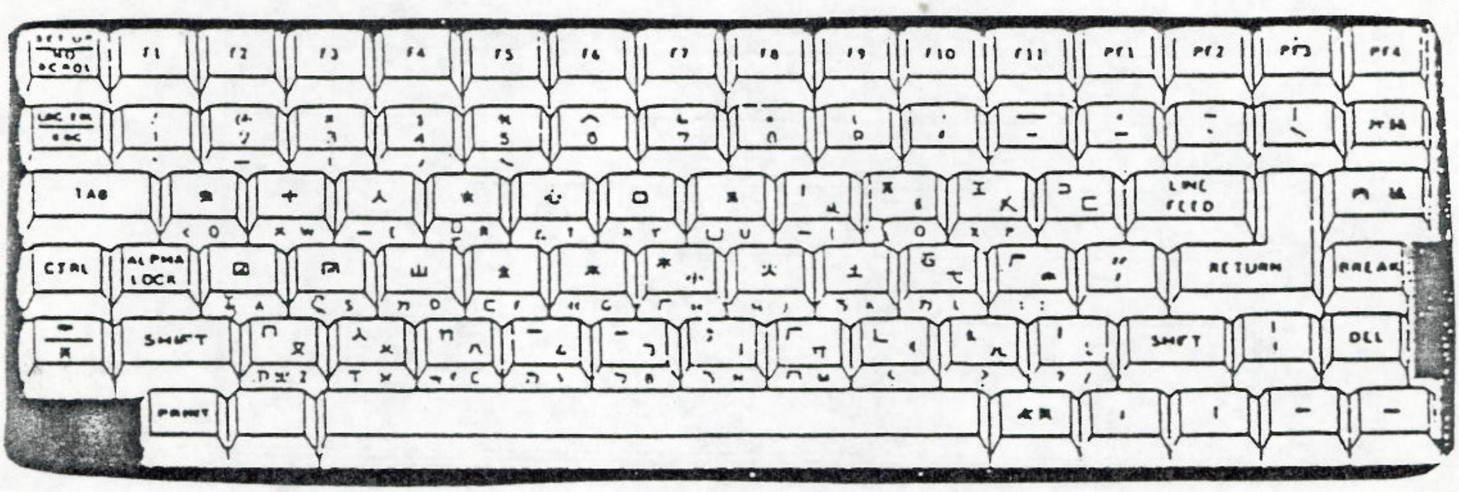
SIMPLEX CODE

Used by at least 7 Taiwanese companies in 1986

字首常用三十字形及其同码输入字形					
字形	编码	同码输入字形	字形	编码	同码输入字形
金	金		口	口	
木	木	禾	三	三	彡
水	水	氵	扌	手	扌
火	火		言	言	
工	工	工工土	三	王	
区	巳	巳巳	一	冠	冖冖
冈	与	风区凶⺈	入	入	
山	山		一	竹	
石	石		廾	草	廾
尸	門	户門門	丷	米	丷丷丷
田	田		勹	夕	夕夕夂
十	十	十扌乂	厂	倉	厂厂厂厂
人	人	亻亻	辶	走	辶
女	女		幺	絲	糸糸糹
心	心	忄	礻	示	礻



Examples : 鈴: 金 丶 ㄌ
燈: 火 一 ㄌ
雄: 十 隹 ㄐ



from: Chang, Ifay and Yu, Wellington. <u>Survey of Chinese input methods: International Conference on Chinese Computing Tutorial Papers</u>. Singapore: University of Singapore, Institute of Systems Science, 1986.

ATTACHMENT 6

4-CORNER CODE

Used by at least 2 Taiwanese companies in 1986
Numeric code 0-9

(表一)四角號碼檢字法

四角號碼檢字法之基本法則

筆畫分為十種，各以號碼代表之如下：　　　　王雲五發明

號碼	筆名	筆形	舉例	說明	注意
0	頭	亠	言主广疒	獨立之點與獨立之橫相結合	0456789各種川由數筆合為一複筆，檢查時遇單筆與複筆並列，應儘量取複筆，如亠作0不作3，十作4不作2，厂作7不作2，⺌作8不作32，小作9不作33。
1	橫	一乚乀	天土地江元風	包括橫、挑與右鈎	
2	垂	丨丿亅	山月千則	包括直撇與左鈎	
3	點	丶乀	宀ネ冖厶之衣	包括點與捺	
4	叉	十乂	草杏皮刈大對	兩筆相交	
5	插	扌	才戈申史	一筆通過兩筆以上	
6	方	口	國鳴四目甲由	四邊齊整之形	
7	角	ㄱ𠃌𠃊厂𠂆	羽門灰陰雪衣學罕	橫與垂相接之處	
8	八	八丷人𠆢	分頁羊余災疾朱午	八字形與其變形	
9	小	小⺌忄	尖糸粦呆惟	小字形與其變形	

標 4199　　王 1010

芬 4422　　主 0010

筆畫號碼歌
(胡 適)
一橫二垂三點捺 點下帶橫變零頭
叉四插五方塊六 七角八八九是小

from: Chang, Ifay and Yu, Wellington. Survey of Chinese input methods: International Conference on Chinese Computing Tutorial Papers. Singapore: University of Singapore, Institute of Systems Science, 1986.

ATTACHMENT 7

STROKE ORDERING CODE

Used by Sinotronics, Hong Kong (developed in Melbourne)
numeric code 1-5

CODE TABLE

1 2 3 4 5

丶 一 丨 丿 乀

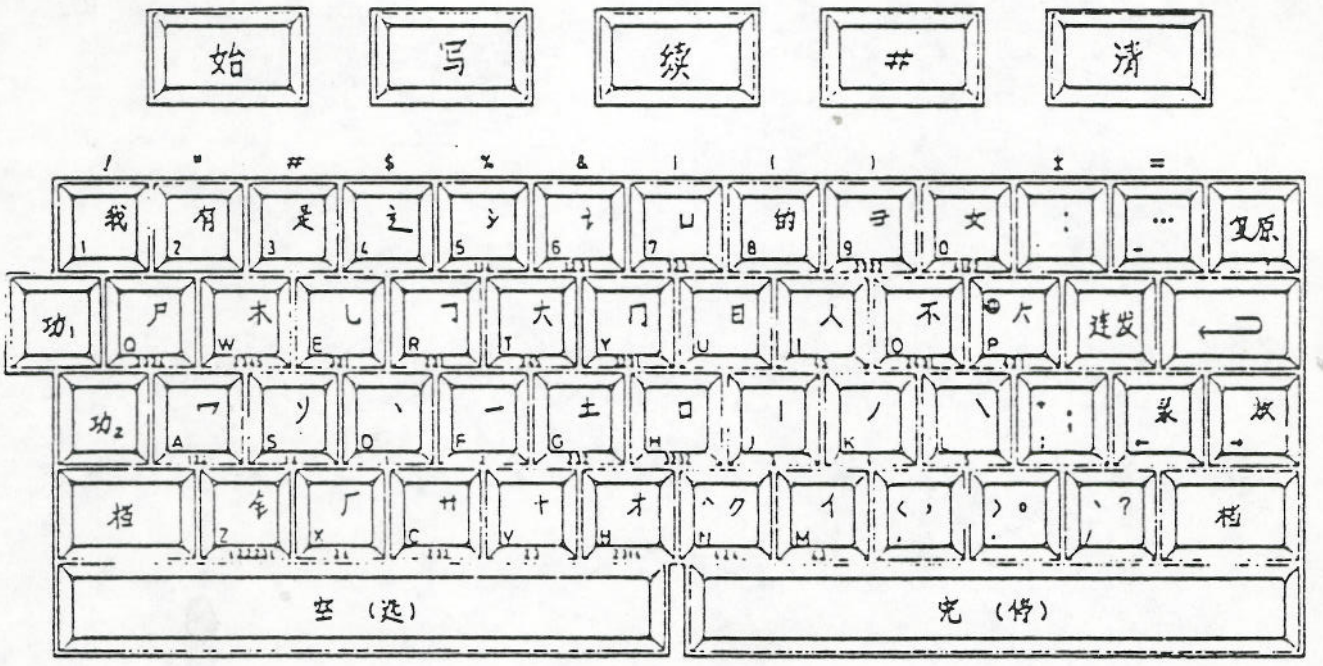
EXAMPLES

口 土 王 工 士

3232 232 2232 232 232

人 八 入 露 琼

45 452 4522 21243 1111 3 232 22321 23232 314



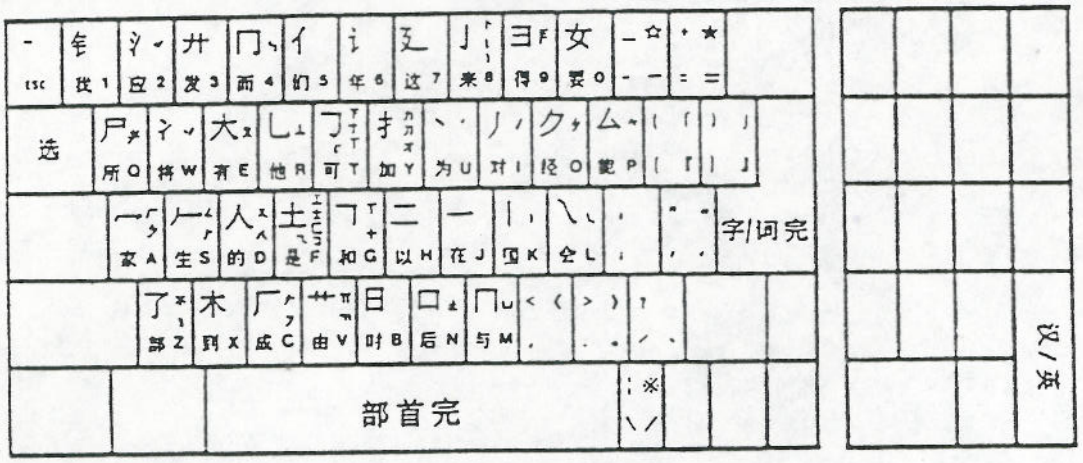
from: Chang, Ifay and Yu, Wellington. Survey of Chinese input methods: International Conference on Chinese Computing Tutorial Papers. Singapore: University of Singapore, Institute of Systems Science, 1986.

ATTACHMENT 8

RADICAL INPUT METHOD (RIM)
(BRUSHWRITER)

Used by Kompac Technology, Singapore
36 radicals

Macintosh Plus



IBM PC

Atari 1040ST