

# Unravelling the Chinese typesetting puzzle

Brian Gaines

Chinese, the oldest recorded language in use today and spoken by one quarter of the world's population, is one of the few challenges left for modern computer photosetting systems.

In September 1978 Monotype decided the time was ripe to tackle the problem and by December demonstrated an embryo system in Hong Kong. A joint development programme with the China Printing Corporation led to a complete pre-press system being developed and installed in Beijing and Shanghai by July 1979. After discussing the problems of Chinese typesetting, the author, who led this remarkable project, describes existing solutions and the Monotype system.



**Brian Gaines**

has been group technical director and deputy chief executive of Monotype Holdings Ltd since its formation in November 1978. He was previously Professor of

computer engineering and chairman of the department of electrical engineering science at the University of Essex and continues to hold a number of honorary academic appointments. He has also previously been technical director of Micro Computer Systems Ltd, a computer manufacturing company, Questel Ltd, a computer bureau, and City Computer Systems Ltd, a computer systems supplier. He holds a first degree in Mathematics and a doctorate in Psychology both from Trinity College, Cambridge.

One by one the languages of the world have been conquered by the modern technology of electronic keyboards, text editors and phototypesetters. But there are a few challenges left, one being Chinese the oldest recorded language in use today. Its commercial importance lies in the quarter of the world's population for whom Chinese is the main language. Its technical difficulty lies mainly in the many thousands of different Chinese characters required and to some extent in the complexity of the characters themselves.

In September 1978 Monotype decided that the time was ripe to tackle the problems of the Chinese language. Microprocessors and digital typography provided an adequate technical base, while China's drive for the *Four Modernisations* created a demand for modern technology. As a company which had supplied China Printing Corporation with equipment for over 60 years, Monotype was in a strong position to market new technology in China. In December 1978 the company put on a demonstration of Chinese typesetting in Hong Kong and agreed a programme of joint development with the China Printing Corporation. Progress was rapid and by July 1979 complete pre-press systems had been developed and installed in Beijing and Shanghai. Keyboards and visual display editors had been manufactured and typefaces of over 5000 characters had been digitised. During the next three months operators were trained and books were set, culminating in major demonstrations to over 5000 representatives of printers in China in October 1979. This article gives the background to this joint achievement of Monotype and the China Printing Corporation.

It is impossible to define precisely the number of characters, or ideograms, in the Chinese language. There are over 60000 ideograms recorded in use during different periods of Chinese history while the modern standard dictionaries used in China list some 13000 to 14000 characters currently in use. For the printing of books a face of some 7000 characters is adequate and for newspapers about 4500 characters. Chinese typewriters provide about 2000 characters available in the type case under the print head and about another 2000 available for insertion as required.

Chairman Mao Zedong in all his writing used a vocabulary of only 3006 characters. There is a major movement in China to simplify the language by restricting it to only 3260 characters but this is a contentious issue. For printing purposes, no matter how many characters are provided there will always be the need for more

through a good 'sorts' facility since specific jobs require access to non standard characters, for example in quoting from an ancient Chinese work. It should also be noted that the calligraphy of Chinese characters was greatly simplified in China after the liberation in a move to aid literacy. A standard form of phonetic romanisation of the Chinese language, Pin Yin, was also introduced and is widely used in China; however it has yet to have any major impact on the printing industry. The direction of the setting of Chinese text was also changed to correspond to the Western format of horizontal, left-to-right reading, rather than the original Chinese vertical setting from right to left. The simplified characters and horizontal setting continue to be used in China, Singapore and Malaysia but the original characters and vertical setting are still used in Hong Kong and Taiwan.

Because of the large number of characters required, Chinese text is still primarily hand-set with hot metal. The operators work within an alcove of type cases containing the characters they need for the job being prepared. The arrangement of the characters and number available is a feat of organisation that minimises the effort in hand picking for the particular job. The configuration required for rapid setting of newspapers can be quite monumental and skilled operators on routine material can achieve speeds of up to 1 000 ideograms an hour. As a rough guide in translating these figures for comparison with Latin languages a three to one ratio has been found appropriate in terms of translated material, that is one Chinese character requires about three English ones on average. Thus a comparable setting rate for English would be about 3 000 characters an hour.

Flat-bed, hand-operated filmsetters made in Shanghai are also in routine use in China mainly for technical book production. The machine shown is manufactured in Shanghai and provides 9 555 characters on a five by seven matrix of glass plates each of which has a 13 by 21 matrix of characters. A turret lens system allows the point size to be adjusted over the range four to 60. Some of the plates contain mathematical and chemical symbols, Latin and Cyrillic alphabets, and so on. They are readily interchanged to provide the particular faces required for specific jobs. The operator moves the main bed around whilst viewing the characters through a magnifier. When the one required is found a lever is moved which engages a ratchet to fix the precise location of the character to be exposed. Skilled operators can achieve rates of the order of 1 000 characters an hour which is comparable with hand-picking type.



These then are the typesetting technologies with which any new approach has to compete. Both the hot-metal and film systems give access to the very large number of characters required to set Chinese; both give the capability of setting Chinese mixed with other languages and technical material; and both give a very high quality of output. The main disadvantages of the two systems are that they are manually operated at a fairly slow rate, require skilled operators with some three years of training to reach full output, and give no facilities for text storage, editing and aids to composition. The electronic phototypesetting techniques that have developed so rapidly in the West and are used so extensively for both book and newspaper production have so far proved unsuitable for Chinese primarily because of the numbers of characters required, but also because of the high status of the calligraphic arts in China which demand the highest quality. It is salutary when examining the speed and effectiveness of hand-setting in China to note how competitive it is with modern technology. In a country where labour costs are low and technical materials are expensive then hot-metal is still a very cost-effective technology.

The initial impetus for work on Chinese typesetting at Monotype came from our realisation that the fourth generation of laser

Flat-bed filmsetter made in Shanghai. (All colour photographs by the author.)



phototypesetters with digital character storage were ideally suited to ideographic languages with large numbers of characters. In the Lasercomp characters are stored in digital form on magnetic discs under computer control. They are best regarded as digital versions of photographic images scanned at a resolution of 1 000 lines an inch. These images may be placed anywhere on the output media, photographic paper or film, at any size from four up to 256 point on any measure up to 100 picas. There are no intrinsic limitations on the numbers of different characters that the Lasercomp can handle. The only practical limitation is the storage capacity of the magnetic disc and since this ranges from 80 Megabytes up to 320 or more this was no obstacle. Even the smallest size can contain some 60 000 Chinese



Hand setting the 6.3m daily run *People's Daily* newspaper from a 3500 character set. Speeds of 1000 characters are achieved an hour – equivalent to 3000 Latin language characters. (Photo : Bruno Barbier).

characters in normal book faces, and the larger sizes of disc allow for a wide range of faces, sizes and languages.

Thus the storage limitations which made previous generations of phototypesetters unsuitable for Chinese no longer applied to the Lasercomp. It was also reasonable to suppose that the laser optical system with its precisely controllable, high-resolution image and lack of halation would give the typographic quality required to compare with hot-metal setting of Chinese. To evaluate the machine in this application 300 Chinese characters were digitised rapidly in the typographic department at Monotype, Salfords, and shown on the machine at the demonstration in Hong Kong in December. The reaction from delegates of the China Printing Corporation to the



1



2

quality was very favourable. Later evaluations by publishers of the books set on the equipment in China confirmed these initial impressions over a wide range of material and type faces.

Whilst the typesetter presented no problem the same was not true for the keyboard technology. In China itself there are over 200 known keyboard designs under development and evaluation and other designs of ideographic keyboards are in use or under development in Japan. Some use a large bank of keys from 700 to 2000 with multiple shifts so that each key represents from four to 12 characters. Others use multiple key depressions for a single character representing: the character in phonetic form; as the sequence of strokes making up the written character; as a set of shapes, or *radicals*, making up the character; and so on. All these different forms of keyboards have their merits and demerits, their proponents and opponents, and there are many hundreds if not thousands of written articles describing them. There have also been a number of attempts to romanise Chinese characters and represent them in a Latin script.

It would take a separate article even to begin to describe the plethora of Chinese keyboards, the relationships between them, the underlying approaches, applications studies, and so on. It is more relevant here to note some of the logic which a commercial



3



4

- 1) Visitors at the exhibition in Beijing – Mr Wang Yi, Chairman of China Printing Corporation and Dr Peter White, Chairman of Monotype.
- 2) Professor S C Loh, inventor of the Loh keyboard, demonstrating a Chinese Telex link in Beijing.
- 3) Telephone link in Shanghai.
- 4) Demonstration placards outside the Xin Hua works.

manufacturer must apply in selecting a suitable keyboard system for a language. The primary criterion is not a technical one, not one of speed, accuracy, speed of learning, or cost of manufacture; it is that of acceptance to the purchasers. Until recently no standard had been adopted in China but in 1978 it became apparent that keyboard system developed by Professor S C Loh at the Chinese University of Hong Kong had a good chance of becoming a standard for a high-speed, multiple-key per character, keyboard in China. Loh is an expert on machine translation of Chinese who has studied the range of approaches to Chinese keyboards in depth. He had designed one that had relatively few keys and yet could cover the full set of characters in current dictionaries. In 1978 he had toured China at the invitation of the government lecturing on keyboard design and his own approach had won a high degree of acceptance.

Hence Monotype took out a licence to manufacture the Loh keyboard and developed a prototype for the demonstration in Hong Kong in December 1978. It was possible to put this into production rapidly as a variation of a standard Monotype keyboard, the LD400, normally used for mathematics and multiple language typesetting. The Loh keyboard has 238 keys arranged on a matrix of 14 rows and 17 columns. This was fitted in the centre of the 17 rows by 26 columns of





Table with 2 columns: Unit Name (e.g., 上海第一织布厂) and Contact Info (e.g., 总机521503).

Table with 2 columns: Unit Name (e.g., 后勤组) and Contact Info (e.g., 539426).

Table with 2 columns: Unit Name (e.g., 革委会) and Contact Info (e.g., 241032).

在联邦德国的日子里

华总理访问西德极其成功

德意志通讯社柏林二十一日专电：西德总理阿登纳在柏林机场迎接中国总理周恩来，这是德中关系史上的重要时刻。

市革委会欢宴泰国国会代表团

本市革委会二十一日晚在锦江酒家设宴招待泰国国会代表团，席间宾主尽欢，气氛热烈。

英国首相麦克米伦先生访沪

英国首相麦克米伦先生于二十一日下午抵达上海，受到市革委会和各界人士的热烈欢迎。



英国首相麦克米伦先生（左二）在上海会见市革委会成员。

国家经委精减文件停办六种简报

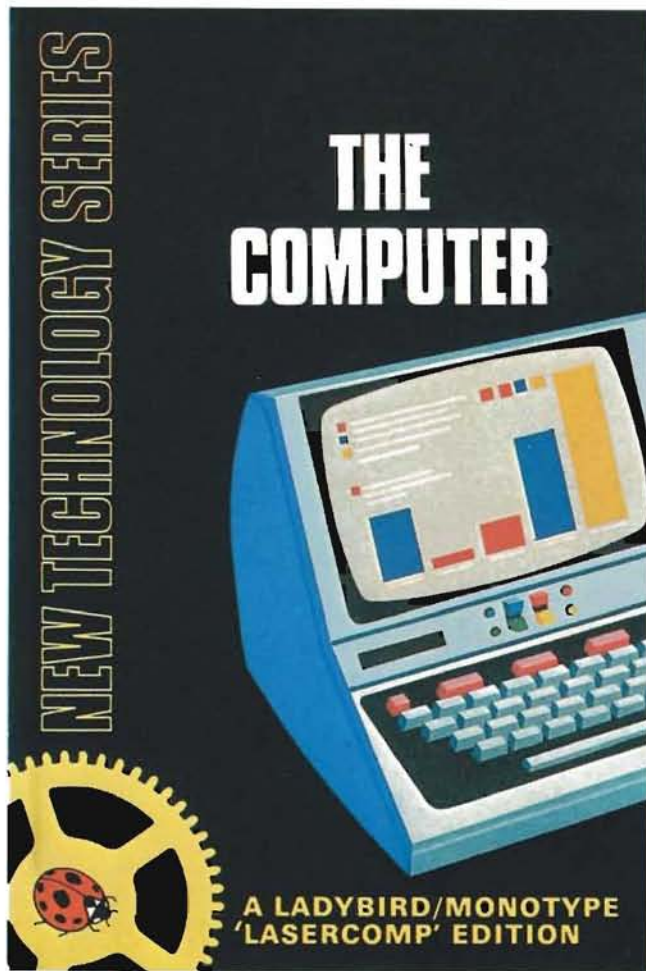
新华社北京二十一日专电：国家经济委员会为精简文件，提高工作效率，决定停办六种不必要的简报。

越南和柬仿脱国际救济来人民

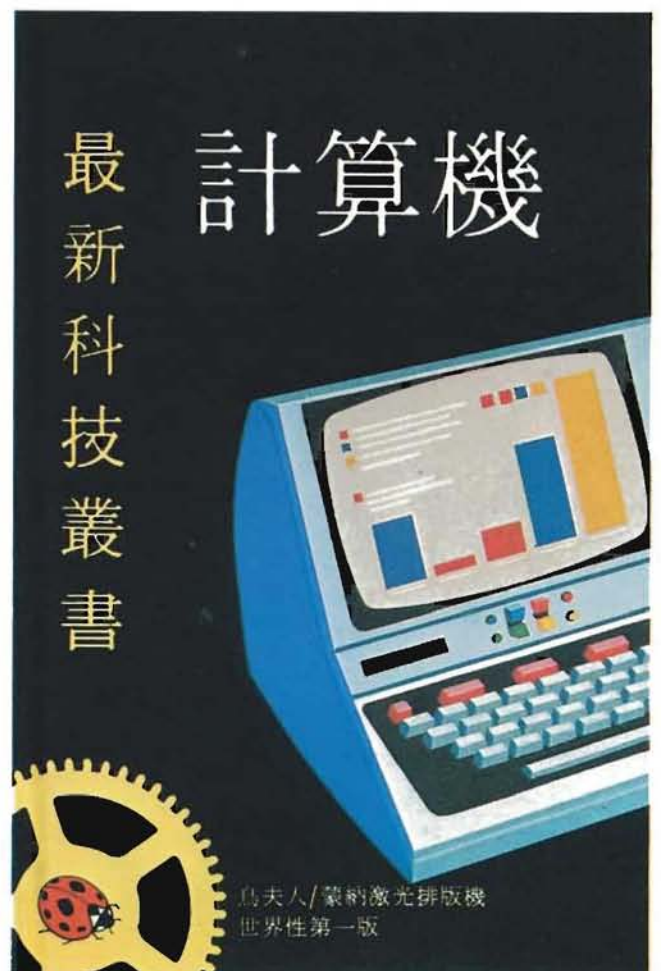
新华社北京二十一日专电：越南和柬埔寨两国政府请求国际救济，以解决国内严重的经济困难。

朝鲜《劳动新闻》评补正熙被枪杀

朝鲜《劳动新闻》二十一日发表评论，严厉谴责南朝鲜当局枪杀补正熙的暴行，认为这是对国际和平的严重挑衅。



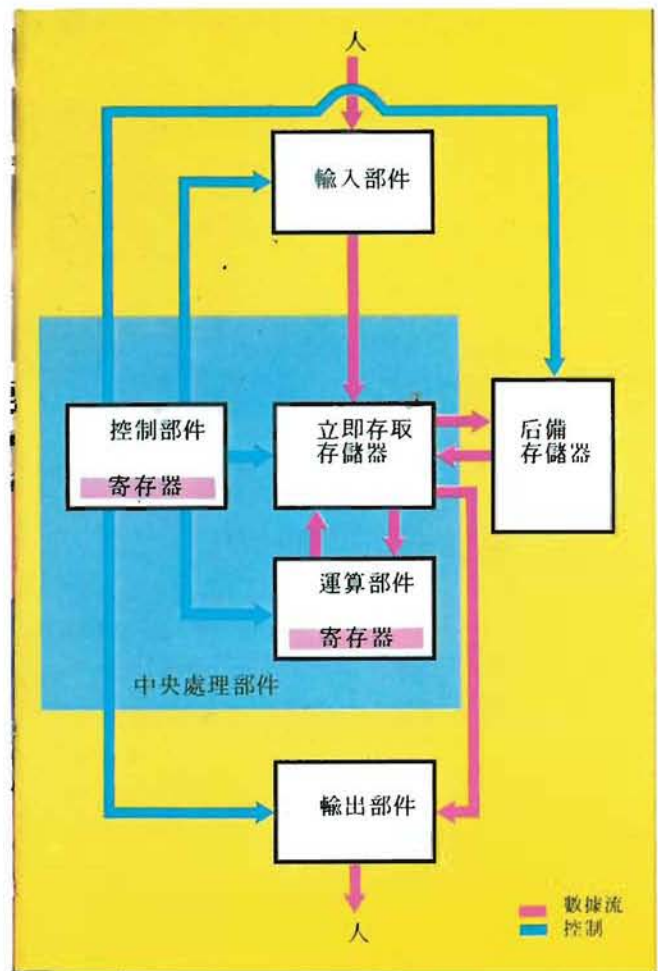
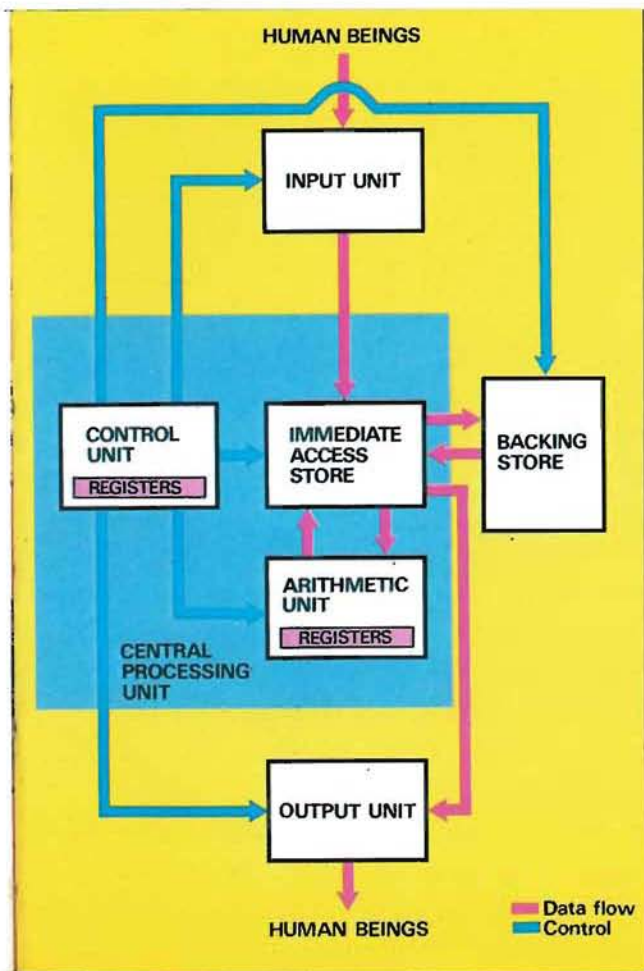
Cover and inside page from Ladybird book on computers printed in English and machine-translated Chinese, set on Lasercomp and used as promotion items.



turn the Loh sequences back into a call for a specific Chinese character within a fount. It is also worth noting that the output from the Monotype Loh keyboard is a set of conventional eight-bit codes which may be communicated through standard serial lines to computers and over wire services.

With the Loh keyboard as a text preparation device and the Monotype Lasercomp as an output device, it remained only to develop a suitable text editing terminal for the Monotype Ideographic Typesetting System (MITS) to be complete. This proved surprisingly easy using modern microprocessor and display technology. Apart from the Loh keyboard the Monotype ideographic editing terminal was manufactured from modular computer components available as standard units. It consists of two Zilog Z80 microprocessors, one with 64 Kbytes of memory used as an editor, and the other with 256 Kbytes of memory used as a character store for the display. Twin 315 Kbytes minifloppy disc units on the editor microprocessor are used for program and text storage respectively, and similar discs may also be used with the fount store to hold the rarer characters beyond the most common 5000. Input to the terminal may be through keyboard, paper tape, floppy disc or serial line, and output may be through paper tape, floppy disc, or serial line.

The editor operates with mixed Chinese, Latin, and Greek text, mathematical and typesetting symbols, in the normal fashion allowing text files to be displayed, searched, modified, split, merged, and so on. The display screen has a resolution of 256 by 256 picture elements



which allows 8 rows of 14 characters to be displayed at a time together with special areas for search strings and status information. It was found possible to represent the Chinese characters adequately with a resolution of 16 rows by 14 columns of dots. Some characters have to be distorted but operators found the screen easy to read at this resolution. A proofing printer is also provided with each editing terminal that operates at the same resolution.

With a keyboard, visual display editing terminal and phototypesetter it was possible to put together for the first time complete phototypesetting systems for Chinese. In the December 1978 demonstrations we had shown only a prototype keyboard working into a Lasercomp with only 300 Chinese characters. For the demonstrations in China a different order of magnitude of system was required. In March 1979 a delegation from the China Printing Corporation led by Mr Wang Yi, the Chairman, visited Monotype's headquarters at Salfords in Surrey, England, and the final arrangements were made. It was agreed that complete book production systems would be installed in the Xin Hua Printing Company in Beijing and also a new printing factory being built in Shanghai by the end of June. Chinese operators would be trained to use and maintain these in time for demonstrations to printers and political leaders from all over China in October 1979 at the time of the celebration of the 30th anniversary of the liberation.

The systems demonstrated are illustrated in the block diagram and consisted of ten keyboards, three editing terminals, two Lasercomps



Cover of Lasercomp-set 12 page text of Chairman Mao's speech to the music workers.

and film processors, and one character digitiser in each of the two centres. These systems make provision for the storage and transmission of text either on paper tape or on floppy disc, or through any combination of the two. In use up to ten operators can key in text directly either to tape or to disc. These tapes or discs may be proofed either in final form through one of the two Lasercomps or as character sequences through the printers on the editing terminals. Tapes or discs may be corrected on one of the three editing terminals and the resultant tape or disc used to produce the final output on paper or film on one of the two Lasercomps.

To ensure that some printed book material was available for the demonstrations Monotype digitised a basic vocabulary of 3260 Chinese characters at Salfords and used this to set a book in Chinese. The material chosen was a new edition of a *Ladybird* book on *The Computer* which was due for publication in 1979. To emphasise the impact of computer technology on text processing the English version of this text was machine translated into Chinese using a program developed by Professor Loh's Machine Translation Research Unit at the Chinese University of Hong Kong. The Chinese text was deliberately left unedited so that the effect of keyboarding a book in English and then typesetting it in both English and Chinese without further human intervention could be seen.

The limited number of characters digitised at Salfords in one type face only was not sufficient for setting books in China, and arrangements were made for Shanghai Printing Research Institute to draw and digitise other faces and a wider range of characters. This institute was responsible for drawing the simplified faces adopted in China after the liberation and had the master drawings. In May 1979 two Monotype digitisers were installed in Shanghai and by August some 5000 characters had been digitised for Lasercomp in the face used for the majority of books in China. Bold and italic versions of this face were also digitised so that complete books could be set and compared directly with the hot-metal equivalents.

By the end of July 1979 complete systems had been installed in Beijing and Shanghai and Chinese keyboard operators commenced training. Six engineers from the China Printing Corporation also came over to the Monotype works at Salfords for training on the maintenance of the equipment. Some twenty Monotype engineers, training staff and demonstrators were also on tours of duty in China during the period from May to October 1979 for periods of two to



向特格雷怪难受,感到死亡就在眼前。但是他还鼓起勇气对强盗讲几句话。

他说:“他们追赶上来了,要捉住你啦,难道你不想对他们说些俏皮话,嘲笑他们一下吗?”

“唉,品特,可怜的品特啊!你将会说我不管怎样,我得穿上皮大氅,不管怎样!”

列那每走一步,向特格雷悲哀地重复一句“不管怎样,不管怎样。”使得列那忍不住,突然也骄傲地重复说:

“不管怎样,是的,不管怎样,你将穿上皮大氅了!”

当他这样自夸狡猾的时候,他的牙齿稍微放松了一些,向特格雷抓住这个机会,挣扎脱身,失去了几根羽毛,却拼命地飞到一根树枝上,在那儿,抖抖翅膀,摇摇身体,说:

“喂!表兄!你皮大氅上的装饰品真硬得很啊!我宁可同你称兄道弟了。我再也不唱歌了,正象我再也不得不睁着一只眼睛来睡觉。”

列那气虎虎地说:“我呀,我再也不能只顾讲话,不顾到嘴里的东西了。”

这时候,一群狗随着农庄的许多仆人奔上来了,列那没有必要把皮大氅送给那些狗,他就远远跑开,真是乘兴而来,败兴而归。

列那上了公鸡的当,在他是一桩奇耻大辱!

### 列那放弃肉食

列那抢劫向特格雷没有成功,可是他对这座花园留下了深刻的印象,因为那里有成群的公鸡和母鸡自由自在地徘徊着。

那一天,由于行动太鲁莽,没有达到目的,怎么就能忘记这批肥嫩的家禽呢。本来是可以拿些鲜美的食物带回马贝渡去,给海梅林和

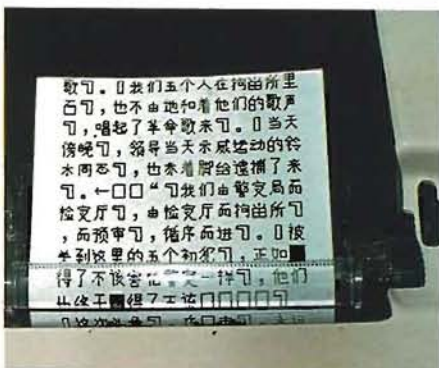
· 67 ·

eight weeks, and the staff of China Printing Corporation and Monotype established a very effective working relationship and many personal friendships. These relationships proved invaluable during the many ups and downs which were inevitable in such a major project undertaken at such speed. When the installation engineers arrived in Shanghai initially the temperature was 90°F, the relative humidity 90 per cent, and the building for the equipment was not yet complete. Air conditioning equipment was rushed out and the schedule maintained but such incidents kept the telex lines between Shanghai and Salfords open and active, and cemented a working relationship between the engineers at both ends which made the whole operation successful and pleasurable.

By the end of August 1979 complete systems were in operation in both Beijing and Shanghai with trained operators and maintenance staff. Experimental production of books then commenced together with a very wide range of other demonstration material such as telephone directories, newspaper pages, complex tabular and mathematical work, Chinese music, and so on. A 155 page book of French fairy stories translated into Chinese was set and printed in Shanghai and a 182 page book in Beijing. A 12 page booklet was also set and printed in Beijing which contained Chairman Mao's famous speech to the music workers in which he calls upon them to 'take that which is best in the West and make it Chinese'; this seemed a particularly apposite handout for visitors to the demonstrations.

Spread from book of French fairy stories which was the first book to be commercially set on the Lasercomp in Shanghai.

激光



- 1) Chinese word for 'Laser'.
- 2) Proofing printer used on editing terminal.
- 3) Loh keyboards in use at Shanghai.
- 4) Close-up of keyboard.
- 5) Block diagram of Monotype Ideographic Typesetting system.
- 6) Character digitiser in use at Shanghai.
- 7) Close-up of digitiser screen.

After the initial training period it was possible to run some checks of the speed and accuracy of the operators using the Loh keyboard. The keyboard operators consisted mainly of girls aged 15 or so in their first jobs after school together with a few experienced typesetters evaluating the system. The trials after some six to 12 weeks of experience showed an average keying rate over a production period of some 1 500 Chinese characters an hour at error rates averaging under one per cent. One operator by the end of this period was achieving an average rate of some 3,000 characters an hour with an error rate of only 0.1 per cent. Thus it was clear that Loh's claims for the keyboard were justified and it had a high productivity with low errors in the hands of a skilled user. It was also clear from the performance of casual visitors who keyed in material and were delighted to see it appear on the screen of the editing terminal display that the keyboard was intrinsically easy to understand.

By October 1979 all the preparations were complete and the main demonstrations then commenced. By that time the two systems had come completely under the control of the staff of the China Printing Corporation and the Monotype technical staff had returned to England. During October, demonstrations of the system in action for book production were given to some 2 500 visitors in Beijing and some 3 000 in Shanghai. The visitors came from all over China and included Vice Premier Wong Chun.

As well as seeing the system in use for book production and being

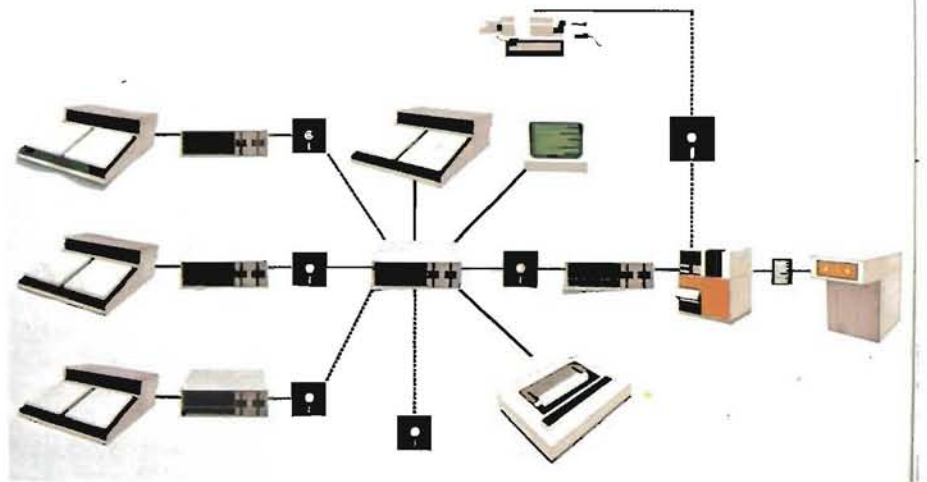


3



4

### Monotype Ideographic Typesetting System (MITS)



5



7



6

able to set up test pieces of their own, the visitors also saw a number of other demonstrations of what could be done in text processing for Chinese using modern electronic and computer technology. One which aroused great interest was the communication of text over a normal telephone line between Beijing and Shanghai. Modems had been installed in the printing works at both ends and it was possible to communicate text from an editing terminal in Beijing to one in Shanghai, or vice versa, and even transmit text directly to the Lasercomps. The most impressive demonstration was one in which the visitor's name was spoken down the phone link, it was switched to data mode and a tape transmitted back which when placed in the Lasercomp produced a personal welcoming text incorporating the visitor's name. Such demonstrations illustrated the capability of the system to overcome some of the problems of physical communications in China; newspaper producers could keyboard text in one location and within minutes have it producing plates at a number of remote locations. They also showed the application of the technology outside the print industry, for example in telex communications. After the demonstrations a link was set up between the Xin Hua works in Beijing and Monotype's Advanced Development Group's laboratory in Cambridge, England, and material transmitted successfully over the international telephone network.

That concludes the story to date. Approval was given by the State Council in China for purchase of the demonstration equipment and for continuation of the joint development of the system with Monotype. Experiments are proceeding in China on the use of photographic materials manufactured there rather than the imported materials used in the demonstrations. Studies are also continuing of the keyboard, operator training, and the maximisation of the overall system productivity. A range of type faces is also being digitised and experiments have commenced on newspaper production. It was clear from the beginning that China could not import all the equipment needed to equip even a fraction of the printing establishments in the China Printing Corporation with its one million employees. Already some essential parts of the Lasercomp have been manufactured in China and the collaboration is intended to extend this to joint manufacture, research and development. The developments reported in this article are only the beginning of a massive programme of activity to modernise not only China's printing industry but its overall information industries.