Gaines, B. R. (1981). Some experience in text processing in the Chinese language. In Sixth ACM European Regional Conference, pp. 301-310. ACM, New York.

SOME EXPERIENCE IN TEXT PROCESSING IN THE CHINESE LANGUAGE

Brian R Gaines

Monotype (China) Ltd., Hong Kong and

G W Information Transfer Systems Ltd., Cirencester, UK.

The Chinese language present many difficulties in text processing. There are some 7,000 characters in routine use and conventional approaches to keyboards, displays and printers are unable to cope with the set required. Yet the language is a very important one since it is in daily use by one quarter of the population of the world. This paper describes a complete phototypesetting system recently developed for use with text in the Chinese and English languages and now in use for book printing in Beijing and Shanghai. Recent work on the application of a similar approach to data processing in Chinese is also outlined.

INTRODUCTION

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 in the many thousands of different Chinese characters required and to some extent in the complexity of the characters themselves.

In the People's Reublic of China itself the drive for Mao Zedong's "Four Modernizations" (of Agriculture, Industry, National Defence and Modern Sciences) has created a demand for modern "information technology". However, so much of this technology has originated from countries using the English language, or at least the Roman alphabet, that it not only requires a knowledge of English for its use but also requires that all information be expressed in Roman characters. For numerical applications of computer systems this does not pose too much of a problem - the specialists, programmers and operators, have to have a working knowledge of English. However, for database and text-processing applications, where the information itself is intrinsically in a non-Roman language, it is extremely problematic to develop any effective systems.

This problem is not peculiar to computers - printing technology has never been well-suited to the Chinese language and there have been a wide variety of attempts to Romanize the script (Seybolt and Chiang 1979). Mao himself is widely quoted for his speech in 1951 when he said, "The written language must be reformed; we must proceed in the direction of phoneticization being taken by all languages of the world", and Zhou Enlai echoed this in 1958, "The immediate tasks in writing reform are simplifying the Chinese characters, spreading the use of the standard vernacular, and determining and spreading the use of phonetic spelling in Chinese". Neither is this a problem peculiar

to Chinese: in Pakistan, for example, some newspapers in Urdu are still produced by a staff of some 60 calligraphers trained to a common handwritten style since there has been until recently no printing technology that can cope with the complexity of the language.

However, in recent years developments in low-cost semiconductor systems have given us new technologies for text and image processing that now make it technically and commercially feasible to develop computer-based systems that operate in any of the languages of the world. Certainly the technical reasons for Romanizing languages such as Chinese have now become far less pressing. Computer technology is a means of making complex tasks simple (although the opposite often seems so!) and offers the possibility of information systems that operate fully in any script. Rather than bend the language to the technology it is now feasible to use the technology to support operations in the language - it does seem reasonable to suppose that a "user friendly" system for use in China should operate in Chinese rather than English.

In September 1978 Monotype decided that the time was ripe to tackle the problems of text capture, editing and phototypesetting in the Chinese language. By June 1979 complete systemsd had been developed and installed in Beijing and Shanghai. This paper gives the background to this development and the technologies used.

TYPESETTING THE CHINESE LANGUAGE

It is impossible to define precisely the number of characters, or ideograms, in the Chinese language. There are over 60,000 ideograms recorded in use during different periods of Chinese history while the modern standard dictionaries used in China list some 13,000 characters currently in use. For the printing of books a face of some 7,000 characters is adequate and for newspapers some 4,500 characters. Chinese typewriters provide about 2,000 characters available in the type case under the print head and about another 2,000 available for insertion as required.

Chairman Mao Zedong in all his writing used a vocabulary of only 3,006 characters. There is a major movement in China to simplify the Chinese language by restricting it to only 3,260 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 nonstandard characters, for example in quoting from an ancient Chinese work. Similarly in database systems it should be noted that the most nonstandard characters are those for personal and place names!

The calligraphy of Chinese characters was greatly simplified in China after the liberation in a move to aid literacy. A standard form of phonetic romanization of the the Chinese language, Pin Yin, was also introduced and is widely used in China for shop and street names; however, it has yet to have any major impact on the printing industry. The direction of 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 are primarily used in China, Singapore and Malaysia, but the original characters and vertical setting are still used in Hong Kong, Taiwan, and, with extensions, in Japan.

Because of the large number of characters required, Chinese text is still primarily hand-set using hot metal techniques in the printing

industry. The operators work within an alcove of type cases containing the characters needed for the text they are preparing. The arrangement of the characters and the number available is a feat of organization that minimizes the effort in hand picking for a particular text. The configuration required for rapid setting of newspapers can be quite monumental with operators literally skating from case to case to achieve high speeds. On routine textual material skilled operators can achieve continuous throughput of up to 1,000 characters 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 translated material, i.e. 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, or just under one a second. This compares unfavourably with the setting rates for skilled keyboard operators with Roman texts.

Flat-bed, hand-operated filmsetters made in Shanghai are also in use in China for technical book production. The machine 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 size of the characters to be varied optically in the range from 4 points to 60 points (there are 72 points to an inch). 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 texts. 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 throughputs of the order of 1,000 characters an hour which is comparable with hand setting.

These then are the typesetting technologies with which any new approach has to compete. Both the hot metal and the 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 slow rate, require skilled operators with some three years training to reach full throughput, and give no facilities for text storage, editing and aids to page layout. The electronic phototypesetting techniques that have been developed so rapidly inthe West and used extensively for book, magazine and newspaper production have so far proved unsuitable for Chinese primarily because of the number of characters required, but also because of the high status of the calligraphic arts in China which demand the highest quality in any printed text. It is salutory also when examining the speed and effectiveness of hand setting in China to note how competetive it is with modern technology. In a country where labour costs are low and technical materials are expensive then the old hot metal techniques are still very cost-effective.

DEVELOPMENT OF A CHINESE PHOTOTYPESETTING SYSTEM

The key subsystems in a phototypesetting system for Chinese are: keyboards for text acquistion and editing; visual displays for editing; printers for proofing; high-quality phototypesetters for final composition. The key additional requirements that the Chinese language places upon these all stem from the number and complexity of the ideographic characters.

The Lasercomp Phototypesetter

The phctotypesetter itself presented no problems. Indeed the initial

impetus for this development came from the realization that in its "lasercomp" laser-based, digital phototypesetter Monotype had available the technology to set Chinese text with the full repertoire of characters and typesfaces required and with the quality also essential. The Lasercomp is essentially a high resolution raster scan plotter in which a Helium-Neon laster is used to form an image on photographic material. The laser beam is one thousanth of an inch in diameter and deflected horizontally by a spinning mirror system. Vertical traverse is through the movement of the film material in synchronization with the mirror. Essentially one can place a dot 1 thou in diameter anywhere on film material 58 picas or 100 picas wide (10 inches or 17 inches wide) and of virtually indefinite length.

The lasercomp is controlled by a Computer Automation lsi4 computer with special microcode for graphics. Character founts are held in digital form on CDC 80 MB discs and there are no intrinsic limitations to the number of different characters the lasercomp can handle. The machines used in China had a capacity of some 60,000 Chinese characters, sufficient to store the range of ideograms required in a number of different faces. To evaluate the quality of the lasercomp in setting Chinese text 300 ideograms were digitized rapidly in the typographic department in the UK and shown on the machine at a demonstration given in Hong Kong in December 1979. The reaction then by delegates from China Printing Corporation was very favourable and later evaluations by publishers of books set on the equipment in China confirmed that the quality of output was at least as good as that produced through hot-metal techniques.

A Chinese Keyboard

Whilst the typesetter presented no problems 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 for ideographic keyboards are in use or under development in Japan. Some use a large bank of keys from 700 to 2,000 or more with multiple shifts so that each key represents from four to twelve characters. Others use multiple key depressions for a single character representing it: 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 keyboard have their merits and demerits, their proponents and opponents, and there are many hundreds if not thousands of papers and articles describing them. This is quite apart from system that use the Romanized forms of the Chinese characters and translate them back to the ideograms.

It would take a separate paper to even begin todescribe the plethora of Chinese keyboards, the underlying approaches, the relationships between them, applications studies, and so on. It is more relevant here to note some of the logic a commercial manufacturer must apply in selecting a suitable keyboard for any language. The primary criterion is not a technical one of speed, accuracy, speed of learning, or cost of manufacture. It is one of acceptance to the purchasers - what is the standard which the users, or the country of use, has adopted or will adopted. For example, the QWERTY keyboard is the de facto standard in the UK and USA despite its severe and well documented deficiencies. Unfortunately no such standard yet exists in China. However, it became apparent in 1978 that a keyboard system developed in the Machine Translation Unit at the Chinese University of Hong Kong had a good chance of at least serving as a model for a high speed multiple key per character keyboard in China. This keyboard had been given much publicity in China during 1978 and China Printing Corporation were interested in giving it practical trials for use in book setting. Monotype developed a version of this keyboard incorporating Roman

characters also together with typographic commands and put it into production rapidly as a variant on a standard large keyboard normally used for mathematics typesetting. The Chinese keyboard in the version developed had 238 keys arranged on a matrix of 14 rows by 17 columns. This was fitted in the centre of the 17 rows by 26 columns of the Monotype LD400 keyboard, and typesetting function keys, Latin and Greek alphabets, and mathematical keys were fitted around it. For ergonomic reasons the central region of the Chinese keyboard was designed as 10 rows of 11 columns each representing one symbol. These are arranged so that the first row has the one-stroke Chinese character for 1, 一, at the centre and consists of single stroke characters or radicals. The second row has the two-stroke Chinese character for 2, 二, at the centre and consists of double-stroke characters or radicals. The remaining rows of the central keyboard follow this same logic. The keys to the left of this central region represent components of characters that occur only on the left hand side of an ideogram; those to the right of it occur only on the right hand side; those above occur only in the top part of an ideogram; and those below it occur only in the bottom part. The keys outside the central region each represent two symbols but there is no need for a shift key to distinguish between them since they can never be used as alternatives for one another.

This form of keyboard is able to represent over 13,000 Chinese characters by sequences of only 238 keys without the use of shift keys and with a layout that is very logical, ergonomically well-organized and easy to learn. The actual sequence of keys to depress for a character is derived from the stroke sequence that would be used to draw it and hence is easy for a person trained in Chinese writing to master. Keyboards have been designed with very few keys that allow ideograms to be entered as their actual stroke sequences. However, the average number of key depressions required then becomes over six and text preparation is unacceptably slow. The coding structure adopted by Monotype may be regarded as an enhancement of such basic schemes in which common key sequences are compressed into single key strokes. The average number of key strokes in standard book work is three with a maximum of nine (research at Shanghai Printing Research Institute has recently improved these figures by reducing the maximum to four).

It is worth emphasizing here that the keyboard essentially gives just an unambiguous key sequence for each character. It does not precisely encode the shape or structure of the character. The sequences produced by the keyboard are decoded by computers elsewhere in the system to identify which Chinese characters they represent. Thus the Lasercomp calls its characters by number from its disc and uses a code structure unrelated to that of the keyboard. This is significant in that the entire system was designed to be keyboard independent in view of the probable future variability of keyboard standards in China.

A Chinese Text Editing Terminal

With a suitable keyboard for text preparation and a suitable phototypesetter for text output, it remained only to develop a text editing terminal for Chinese for the Monotype Ideographic Typesetting System (MITS) to be complete. This proved surprisingly easy using modern microprocessor and display technology. The Monotype ideographic editing terminal was manufactured from modular computer components available as standard units. It consists of two Zylog Z80 microprocessors, one with 64 Kbytes of store used as a text editor, and the other with 256 Kbytes of store used as a character store for the display. Twin 315 Kbyte minifloppy disc units on the editor unit are used for program and data storage respectively, and a similar unit is used on the character store to hold the rarer characters beyond the

paper tape, minifloppy disc or serial line, and output may be through paper tape, minifloppy disc, or serial line.

The editor operates with 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 operator interaction, search strings, status information and error messages (all of which are in Chinese). 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 and this consists essentially of an 8-wire matrix printer driven in the "graphics" mode where the wires are directly controlled from the same character store used to drive the display. The standard CPM operating system was used for the Z80s and the editor programs were written in a mixture of macro assembler and C. Modem communication protocols were also written for the editing computers enabling text to be transmitted over telephone links with full error checking and correction.

MONOTYPE IDEOGRAPHIC TYPESETTING INSTALLATIONS IN CHINA

With the keyboard, visual display editing terminal and proofing printer, and phototypesetter it was possible to put together for the first time complete phototypesetting systems for Chinese. In the December 1978 demonstrations in Hong Kong, 3 months from the start of the project, we had shown only a prototype keyboard working into a Lasercomp with only 300 characters. These demonstrations led to invitations for full-scale demonstrations of book production in China the following year and for these a different order of magnitude of demonstration was required. By June 1979 two systems had been installed in China, one in the Xin Hua printing works in Beijing and the other in a newly built printing factory in Shanghai. These systems each consisted of ten keyboards, three editing terminals, two lasercomps, one character digitizer and one film processor.

Each of the two installations was suitable for complete book production. Each had provision for the preparation, 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 could key in text directly to tape or disc. The output could be proofed through the line printers or directly through the Lasercomps. Tapes or discs could be corrected from the proofs or screens at one of the three editing terminals and the resultant tape or disc used to produce the final output on paper or film at one of the two lasercomps.

To ensure that some printed book material was available as early as possible for demonstrations in China Monotype digitized a basic vocabulary of 3,260 characters in the UK and used this to set a book in Chinese. The material chosen was a new edition of the Ladybird book on "The Computer" which was due for publication in 1979. This was a particularly appropriate choice not only for its content but also because the wealth of diagrams and flow charts showed up the advantages of the Lasercomp in complex typesetting.

The limited number of characters digitized in one type face in the UK was not sufficient for setting books in China, and arrangements were made for the Shanghai Printing Research Institute to draw and digitize

other faces and a far 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 character digitizers were installed in Shanghai and by August some 5,000 characters had been digitized for the Lasercomp in the face used for the majority of books in China. Bold and italic versions of this face were also digitized so that complete books could be set and compared directly with hot metal equivalents.

By the end of July 1979 the systems were in full operation and Chinese keyboard and system operators were being trained. Six engineers from the Chinese Printing Corporation also came over to the Monotype works in the UK for 6 weeks training on the maintenance of the system. 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 eight weeks. Conditions in China were difficult at times - when the installation engineers arrived in Shanghai the temperature was 90 f, the relative humidity 90%, and the building for the equipment not yet completed - air conditioning equipment was rushed out and the schedule maintained. Such incidents kept the telex lines between China and the UK open and active, and cemented a working relationship between the engineers at both ends which nade the whole operation successful and pleasurable.

By the end of August 1979 experimental production of books had commenced together with a very wide range of other demonstration material such as newspaper pages, complex tabular and mathematical work, Chinese music, and so on. A 155 page book of French fairy stories translated in Chinese was set 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 to the many thousands of visitors who flocked to the full system demonstrations in October and November of 1979.

Apart from seeing the book production systems these visitors also had the opportunity to see how modern information technology, "made Chinese", might be used for communications and database systems in China. One demonstration which aroused great interest was the communication of text over normal telephone lines between Beijing and Shanghai. Modems operating at 1200 baud had been installed in the printing works at each end 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. Visitors enjoyed speaking their names down the phone link and seeing it then turned to data transmission to produce a personal welcome message on the Lasercomp keyboarded at the remote location.

An important practical application of such communication is in the Chinese newspaper industry where newspapers producers could now keyboard text at one location and, within minutes, be producing printing plates at a number of remote locations. This is particularly significant in China which is a vast territory with difficult physical communications. Such demonstrations showed also that telex communication in Chinese was now feasible. After the demonstrations in November a telephone 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.
The systems in Beijing and Shanghai were purchased after the demonstrations and further orders placed for new systems to be

developed including a cluster edit terminal around a central database capable of driving the Lasercomps directly. It is fascinating to note that the Chinese have moved directly in a very short time from the use of the "zeroth" generation of hot metal printing equipment to the "fourth" generation of digital laser phototypesetters without ever using interveing generations of filmsetters, CRT phototypesetters, and so on.

FURTHER DEVELOPMENTS IN IDIOGRAPHIC INFORMATION TECHNOLOGY

The demonstrations of the use of the Chinese language in the printing industry described above may be taken in a far wider context to show that it is now possible to develop "Information Industries" around languages other than those based on Roman type faces. The significance of this in the spread of computer technology to the majority of the world's population who do not use Roman character sets cannot be overestimated. Clearly some such languages have already been tackled, Arabic in particular. However many of the languages of the Indian and Asian continents have been regarded as beyond computer use currently. This is clearly no longer so. It is now feasible to design "Universal Language" systems that operate in all the languages of the world and whose cost is not very much greater than those operating in the English language alone. The print industry application detailed in this paper is at the summit of difficulty in the range of material, complexity of keyboarding, and quality of output required. Many other equally important applications are very much simpler. In this final section I will outline a number of others under development by G W Information Transfer Systems (GWITS).

Telex

For any language the only requirements for a telex system are a keyboard and a printer interfaced together through a serial communications link. The Monotype keyboard for Chinese was designed with an RS232 output for convenience of interfacing to computers and hence was intrinsically capable of operating over any serial communications link. Telex systems operating in any language of the world, in which when one types in Chinese it comes out as Chinese, in Arabic as Arabic, and so on, are now technically feasible. The commercial and political significance of these may be noted by the fact that currently messages to the Chinese Ambassador in London have to be translated into English for telex transmission and then re-translated into Chinese on receipt. The laboriousness of such procedures and the scope for error is restrictive of all forms of communication.

Typewriters

Clearly similar considerations as to telex apply to typewriters. In general the output has to be of higher quality but low-cost ink-jet printers and variable pitch matrix printers make the printing of arbitrary characters at correspondence quality technically feasible on a cost-effective basis.

Word Processing

The editing terminal in the MITS system is a basic word processor. In any word processor only the keyboard, screen and printer have to be extended to cope with a variety of languages. The text processing software can remain the same for all applications.

Computer Data Entry

The telex and word processing systems already described give a suitable technology for data entry and retrieval from computer systems. In GWITS studies of Chinese database systems we have used the microprocessor in the word processing system to encode fields of Chinese characters into ASCII sequences that appear as character strings to standard datasbase software. Similarly it has been possible to recode the messages from the computer supporting the database as equivalent Chinese messages. This recoding is by no means language "translation" in the full sense of the word but just a table look up requiring knowledge of the system generating the messages. Since this system is a computer, however, the 'language' used is simple, formalized and algorithmically generated so that it is generally easy to recode it.
Thus the modification of the input system of a database designed for use with Latin languages to operate with any arbitrary language whatsoever is straightforward and very cost-effective in comparison with re-writing the database software itself. The technique is essentially one of syntactic signal conditioning in the terminal equipment. In cost terms once the terminal has the keyboard, display and printer capabilities, the marginal cost of the code conversion is only a few percent. Thus the "intelligent" terminal can display its intelligence in languages other than those based on Roman characters.

Computer Programming

It is natural to consider the extension of the data entry techniques described above to programming languages. If one contemplates the problems of programming in English and then adds to these the problems of using a foreign language in an entirely different script then the difficulty of developing indigenous information industries in Asia and Africa can be appreciated. It would be attractive to use syntactic signal conditioning as for databases to recode Basic, Cobol, and so on, programs in Chinese into a semantically equivalent English program that could be compiled or interpreted using standard Western software.

A detailed study of an idiographic front-end processor for accessing English computer systems (Witten & Ng 1979) was undertaken for GWITS and showed that, whilst most language constructs could be translated directly as for the database, a few posed major difficulties. The problems mainly arose with string data where the length and position of items within a character string had no direct relation after the translation. This forces some form of semantic preprocessing for operations involving string manipulation. However it is certainly possible to offer substantial language sub-sets, sufficient for many applications, through the simple syntactic recoding.These studies are continuing with a few to offering a range of computer "languages" operating in a range of actual languages other than English.

Computer Systems

The editing terminal described using Z80's is basically a "personal computer" and there are no problems in making its data processing capabilities also available in Chinese or other idiographic languages. Such a development could be of great importance to third world countries where the use of small computers in light industry, education, communications, and so on, is probably of far greater significance than the availability of very large computing systems. There are no intrinsic problems in making small computer systems that operate completely in such languages as Chinese, Japanese, Hindi and Devanagari.

CONCLUSIONS

This paper has described in detail a project wherby a complete phototypesetting system for Chinese was developed very rapidly and is now in use in Beijing and Shanghai for book production. Later studies of the extension of the approach taken to a wider range of information industry applications, databases, programming, and computer systems are also outlined.

It is suggested that the breaking of the dependence of computer systems upon the English language and Roman character sets will be a major factor in the future development of computer systems. The market for computer systems has been one classically where they are developed in the USA, saturate that market, are exported to Europe, saturate that market, and are then sold to the rest of the world. This has been effective for the numerical applications of computers that have been most significant in the past. However the shift of emphasis to interactive and database applications also emphasizes the requirement for systems to operate in the language of the users. This paper has shown that the technology is now available to make this possible; it remains only to develop it fully.

ACKNOWLEDGEMENTS

A project such as that described in this paper involves the efforts of so many people that it is difficult to be fair in acknowledging particular contributions. The original stimulus for the project came from Dr Peter White, Chairman of Monotype, who had the foresight and imagination to see what could be achieved in China. His vision was matched by that of Mr. Wang Yi, Chairman of China Printing Corporation, who set out courageously to explore the unknown; in particular, we are grateful to him and his staff for their generous hospitality to all the Monotype engineers in China. The Chinese University of Hong Kong provided the knowledge and consultancy that went into the keyboard development, and we are grateful to Mr.Kong for his meticulous work on the keyboard design, and to Professor Loh, the Head of the Machine Translation Unit, for his help in negotiations in China. I will not mention individual names of those who worked on this project within the Monotype organization but just note that many people played essential roles that involved efforts way beyond that normally expected. It has been an exciting project that captured all of our imaginations. I am also grateful to Dr Ian Witten for many stimulating discussions of idiographic computing systems.

REFERENCES

Seybolt, P.J. & Chiang, G.K. (1979) Language Reform in China. M.E.Sharpe, New York.

Witten, I.H. & Ng, Y.H. (1979) An idiographic-language front-end processor for accessing English computer systems, Final Report of a Research Project Sponsored by GWITS, Department of Electrical Engineering Science, Essex University, UK.