Asiagraphics System for Ideographic Automation

C. S. George
I. Introduction

The Asiagraphics System design began a decade ago when the computer-based word processor for Western languages was providing a significant increase in efficiency over the non-intelligent office typewriter. The manual typewriter for ideographs was, and remains, a slow and clumsy device even in the hands of an expert.

In designing the Asiagraphics system, we asked ourselves the following questions: What tasks will be involved in language processing for ideographic scripts? What will the environment be? And, most important, who will be using these systems?

There is little doubt that in the future more and more tasks will be influenced by computer technology. In fact, the microprocessor will probably become more generally used than the fractional horsepower motor is today. At present language processing, word processing, data processing, telex, telegram, electronic mail, and typesetting tasks have been automated for Western languages. In the future automation not only will become more widespread but also very likely will include nearly every aspect of information processing and manipulation.

In the workplace of the future, not only the professional typist will process language; nearly everyone in a business will be involved in automated language processing to some extent. A shipping clerk will use a terminal to check and print out a customer's address. A salesperson will compose a letter to a prospective client on a word processor. Designers will use a CAD (Computer Aided Design) device to speed circuit board design. And even the chief executive of a company will have access to information through automation. At a terminal in his office, he may press a few keys and get up-to-the-minute information on sales, production, personnel, etc. In the future at home as well as at work, almost everyone will have occasion to use automated language processing.

For our design of a system for automation of ideographic scripts, we learn from this imaginary excursion into the future that our system must be designed with a wide variety of users in mind. The professional typist as well as the once-a-week user must be taken into consideration. This variety of users bring with them a wide spectrum of requirements and skills, which we must heed carefully in our design.

We started our work with a very careful look at our users. Who are they and what are their needs? Let us take the professional typist and the chief executive as representatives of two extremes of usage—the dedicated and the occasional user. The professional typist needs an easy-to-use system capable of high speed operation. The occasional user needs a system which is easy to learn and use.
II. System Design

The Asiagraphics system is an overall graded approach to the solution of character selection. By graded approach we mean a single solution with multiple levels of access. It is designed with the professional typist as well as the occasional user in mind. For the professional typist, it is capable of high speed operation. Current operating speeds exceed sixty characters per minutes, and 100 characters per minute should not be uncommon. This high speed is made possible, in part, by the support of touch-typing and by the low average number of keystrokes per character (about four). Since there are very few linguistic rules (about two typewritten pages), and since the system uses much information already known to the operator, the fundamentals of the system may be grasped in a few minutes. Ease of learning is very important for the occasional user.

The Asiagraphics system uses unmodified hardware, available from hundreds of dealers throughout the world. The software can run with some modification on most systems capable of bit-mapped graphics. In addition, following our design requirements, our descriptors each access a unique character, and they are readable by both people and machines. Finally, since we may use our descriptor strings internally for storage and processing, to convert existing Western language applications software to ideographic versions is quite straightforward. As for the character storage capacity, over 50,000 different characters can be processed by the program.

III. Input method

The input method is phonetic-based. It is based on the pronunciation of the character, using either the phonetic system of Wade-Giles and pin-yin in Chinese, and either katakana or modified Hepburn romanization in Japanese and hangul or McCune-Reischauer romanization in Korean.

What is typed on the keyboard to select a specific character consists of two main parts:

a. The pronunciation, as indicated above;

b. The radical family, a single keystroke indicating one of Asiagraphic's special fifty-two radical families. (See the next paragraph on the radical families).

Following these keystrokes, the spacebar is struck and the desired character appears immediately in the screen. For example, if we want to type the Chinese character 女 we should do the following:

a. Type its pronunciation: RONG;
b. Type its tone: " / " (i.e. second tone: rising);
c. Type its radical: "B" (i.e. "BAO");
d. Press the spacebar; and the character 女 will be displayed on the screen immediately.

An extra keystroke is derived from the character itself, that portion of the character called the radical. There are 214 traditional radicals in Chinese, and many dictionaries organize the characters by radicals for identification. Asiagraphics has compressed the 214 radicals into a list of 52 radical
families and assigned one key to each. (A-Z upper case = 26, a-z lower case = 26; 26 + 26 = 52.) This list groups the radicals with the same pronunciation into the same radical family, and so it is easy to learn and easy to apply. For example, both 甲 and 乙 are in the radical family M (i.e. the short form for MU).

The above solution for specification of characters among homophones is similar to the menu solution, where a menu of all the characters with the same pronunciation will be displayed and another keystroke will be needed for final selection. However, in Asiagraphics no menu need be consulted, and you will be able to get the specific character with the unique descriptor of its own.

Based on the fundamental methodology described above, Asiagraphics provides an easy to use entry system and an upward-compatible path to allow an operator to exceed 100 characters per minute.

IV. Special features

1. Standard English keyboard is used, instead of a specially designed, non-standard keyboard;
2. Mixture of roman and CJK characters within the same line with proper spacing of each script;
3. New character generation program, which is available for the creation of new characters to add to the character inventory at any time, even in the midst of text entry. Moreover, it can immediately be processed in the text. This means that one is never without a needed character. If a character is not in the database, it can be added interactively and used immediately.
4. Word processing functions
Many of the standard word-processing functions are available. Functions such as character replace, delete, insert, back-space, page scroll, file storage, and retrieval, etc. are all available depending on hardware capabilities. They provide the users high efficiency in handling office and business work.
5. Data processing functions
Ideographic characters may be used in all data processing functions by using Asiagraphics software. Programs for process control, accounting, database management, and other computer operations are handled as with Western languages.
6. Telex communication
Asiagraphics has the capability of sending and receiving telex messages written in ideographic characters instead of roman letters. This greatly improves communication among Asian language users.
7. Off-line operation
Since the Asiagraphics input code is both human- and machine-readable, text may be generated and edited off-line, using a non-computer, standard typewriter. The text can then be machine read by optical character recognition machines (OCR) for subsequent computer processing.

V. Contributions to automation of the on-line cataloguing of CJK materials

In October 1983 OCLC and Asiagraphics entered into negotiations leading towards the development of a CJK library support package. This program will allow librarians to catalog, search and print CJK bibliographical materials in vernacular scripts.
Following is a sample on-line session with OCLC. This shows a library usage of the Asiagraphics System for the on-line cataloguing of Asian materials.

Screen 1 of 2
VZI
OCLC: 9780182 Rec stat: n Entrd: 830805 Used: 830810
Type: a Bib lvl: m Govt pub: Lang: chi Source: d Illus:
Repr: Enc lvl: I Conf pub: 0 Ctry: cc Dat tp: s M/F/B:10
Indx: 0 Mod rec: Festschr: 0 Cont:
Desc: a Int lvl: Dates: 1951

1 010
2 040 VZI $c VZI
3 049 VZII [China] [c.1] [c.2] [c.3] [c.4]
4 090 B09118.9 $b .C45
5 090 $b
6 100 00 Ching-hsiu, $c Shih
7 100 00 静修，$c 释
8 245 10 Chiao kuan kang tsung k'e shih/ $c Ching-hsiu [Shih] shu
9 245 10 教规纲宗科释／$c 静修 [释] 書
10 250 Tsai pan
11 260 0 Shanghai: $b Hung hua she, $c 1951
12 300 190p. (on double leaves): $b diagr.; $c 21 cm.
13 590 Ngian Collection
14 650 0 T'ien t'ai (Sect) $x Doctrines
15 650 0 T'ien t'ai (Sect) $x Criticism, interpretation, etc.
16 700 01 Chih-hsü, $c Shih, $d 1599-1655. $t Chiao kuan kang tsung
17 700 01 智旭，$c 释，$d 1599-1655. $t 教规纲宗