

**Asia and Pacific Regional Seminar on
Information Technology for Newspaper Publishing**

Organized in cooperation with UNESCO

(11 -14 April 1995, Connemara Hotel, Madras)

ASIAN LANGUAGES AND INFORMATION TECHNOLOGY

A SUMMARY OF ISSUES
FOR NEWSPAPER PUBLISHING

Prepared by

John Clews

CONTENTS

Foreword. Acknowledgements	4
Part 1 - Language and technology background	5
1. The language background to Asian publishing	5
1.1 East Asia - Chinese influences	
1.2 West Asia and Europe - Phoenician influences	
1.3 Europe and America - Christian influences	
1.4 West Asia, South Asia and Central Asia - Islamic influences	
1.5 South and Southeast Asia - Hindu and Buddhist influences	
2. Language planning in modern times	10
2.1 East Asia	
2.2 South Asia	
2.3 West Asia	
2.4 Southeast Asia and Oceania	
2.5 Central Asia	
3. Technological influences on publishing in Asia and elsewhere	15
3.1 The first publishing revolution - movable type	
3.2 The second publishing revolution - industrialization	
3.3 The third publishing revolution - computerization	
Part 2 - Key Language and technology issues	17
4. Computers in the newspaper industry	17
4.1 Mainframes and minicomputers versus proprietary systems	
4.2 Microcomputers - the Macintosh	
4.3 Microcomputers - the PC	
4.4 Systems Integration	
4.5 Future concerns: technological issues which affect publishers	
5. Keyboards	22.
5.1 Keyboard shapes	
5.2 National layouts ("typewriter keyboards")	
5.3 Phonetic layouts ("QWERTY keyboards")	
6. UNICODE and other character coding issues	26
6.1 Limitations of UNICODE	
6.2 Increased costs	
6.3 Range of languages	
6.4 Ways around the disadvantages of UNICODE	
7. Telecommunications and publishing	29
7.1 Telecommunications in traditional publishing processes	
7.2 Telecommunications and new information services	
7.3 A case in point: telecommunications in South Asia	
7.4 A case in point: telecommunications in Central Asia	
7.5 General barriers to multilingual telecommunications	

8. Printing issues	32
8.1 Technical processes	
8.2 A case in point: kerning	
9. Other issues	34
9.1 Sorting	
9.2 Retrieval	
9.3 Machine translation	
9.4 Machine transliteration	
10. Looking beyond the basic issues	36
Part 3 - Synthesis of basic issues and recommendations	37
11. Synthesis of basic issues	37
11.1 Overview	
11.2 Prime issues concerning languages	
11.3 Prime issues concerning computers	
11.4 Standardization requirements and realities;	
11.5 Conclusions on the impact of technology in Asian publishing	
12. Recommendations for action	39
12.1 Government action	
12.2 Regional cooperation: assessment of existing organizations	
12.3 International cooperation: possible roles for Unesco and others	
12.4 Action by the private sector	
12.5 Action by individuals	
References	42

FOREWORD

Computerization is now so much a part of modern newspaper publishing, and computerised equipment is used not just by printers. but increasingly' by reporters and editors. This paper provides a general overview of the use of computers in publishing activities and the particular considerations relating to Asian scripts. It is aimed both at those with some technical knowledge of the use of computers in publishing. and also at those who may have little experience with these issues. but who may at some time need to have appropriate information and guidance in this area in order to make decisions for the future.

In Part 1. the historic language background to Asian languages. as well as language planning issues in all parts of Asia and the Pacific are explored. along with the basic technological influences on publishing in Asia and elsewhere, in particular movable type, mechanization, and computerization.

Part 2 examines key issues relating to language and technology, such as how publishing can best use computers, and specific issues such as keyboards. language coding, telecommunications, printing, sorting and retrieval for archiving.

Part 3 makes various recommendations to publishers. governments and international organizations as to how they could improve their use of the new technology in Asian language publishing.

Acknowledgements

Acknowledgements are due to several people, both in Europe and in Asia, who have given of their time in passing on their expertise in this area. both during visits to see them in London, and during in telephone conversations"

My particular thanks go to Mr. Justin Arundale (formerly librarian at The Independent newspaper, and now Brighton University). Mr. Z. H. Zhang (East Word and South China Printing Co.), Dr. Ali Assam and Mr. Hani Lazim (Diwan), Mr. Imad Bachir (Research Editor. Al-Hayat newspaper), Mr. Aurobind Patel (typographer). Mrs. Fiona Ross (Typographic Researcher with Linotype-Hell, Cheltenham), Mr. Misha Wolf (Reuters), Mr. Maurice Bauhahn (consultant with UNICEF, developer of Macintosh-based word processor for the Cambodian Government). Dr. Nguyen Ngoc Tri (Curator. Vietnamese Collections. British Library), staff of the Central Asia Research Centre (School of Oriental and African Studies. University of London), Mr. Preeth Chengappa (Centre for the Development of Advanced Computing, Pune), Professor Nyima Trashhi (Tibet University Computer Department. Lhasa), Professor Wushur Slamw (Xinjiang University, Department of Electronics and Information, Urumqi), Professor M. Enghebatu (Mongolia Language Institute, Inner Mongolia University, Huhhot), Dr. Gerard Gautier (Kaohsiung), Professor Kim Kyongsok (Seoul) and Mr. Shigenobu Katoh (Chief Senior Researcher, Toppan Printing Co., Ltd, Tokyo).

Part 1 - Language and technology background

Part 1 (sections 1-3) provides a background to the technical issues which are discussed in greater detail in part 2 (sections 4-9). Most of the technical issues are derived from the way mankind has evolved language, and more recently evolved computers - both are a part of human evolution.

1. The language background to Asian publishing

Asian scripts have influenced all other scripts worldwide: Asia has been the cradle of writing. and therefore of civilization. Scripts have tended to follow religious and other cultural influences over long periods of time, and in fact all scripts in current use today can trace their ancestry to parts of Asia.

Within the last century. government policies all over the world have determined and shaped some aspects of language use. and standardised things like spelling and the range of characters that can be used. The use of computers has also had a standardizing effect. in determining the range of characters used. and their sorting order.

Although the number of languages in use worldwide still runs into the thousands. the number of scripts currently used to write them is only around two dozen. To simplify things further. it is rarely realised that there are in fact only three basic types of script - good ideas travel and are widely adopted. The three types of scripts grew up in three areas, each largely separated by mountainous and desert areas:

- in East Asia. where Chinese ideographic script has had a major influence:
Seal script (see Figure 1) was one of the earliest ideographic scripts used in China, and an ancestor of the current script:
- in South Asia, in and around the Indian Subcontinent. influenced by Brahmi script: and
- in West Asia and around the Mediterranean, influenced by Phoenician script.

These basic scripts (Figure 1) have a major influence on the scripts we use today, and understanding these three base scripts can give us a major insight into all other scripts and languages we may have to come across. All of them tended to spread with particular religions-and other cultural influences. These three script families are examined below: more detail can be found in other publications.

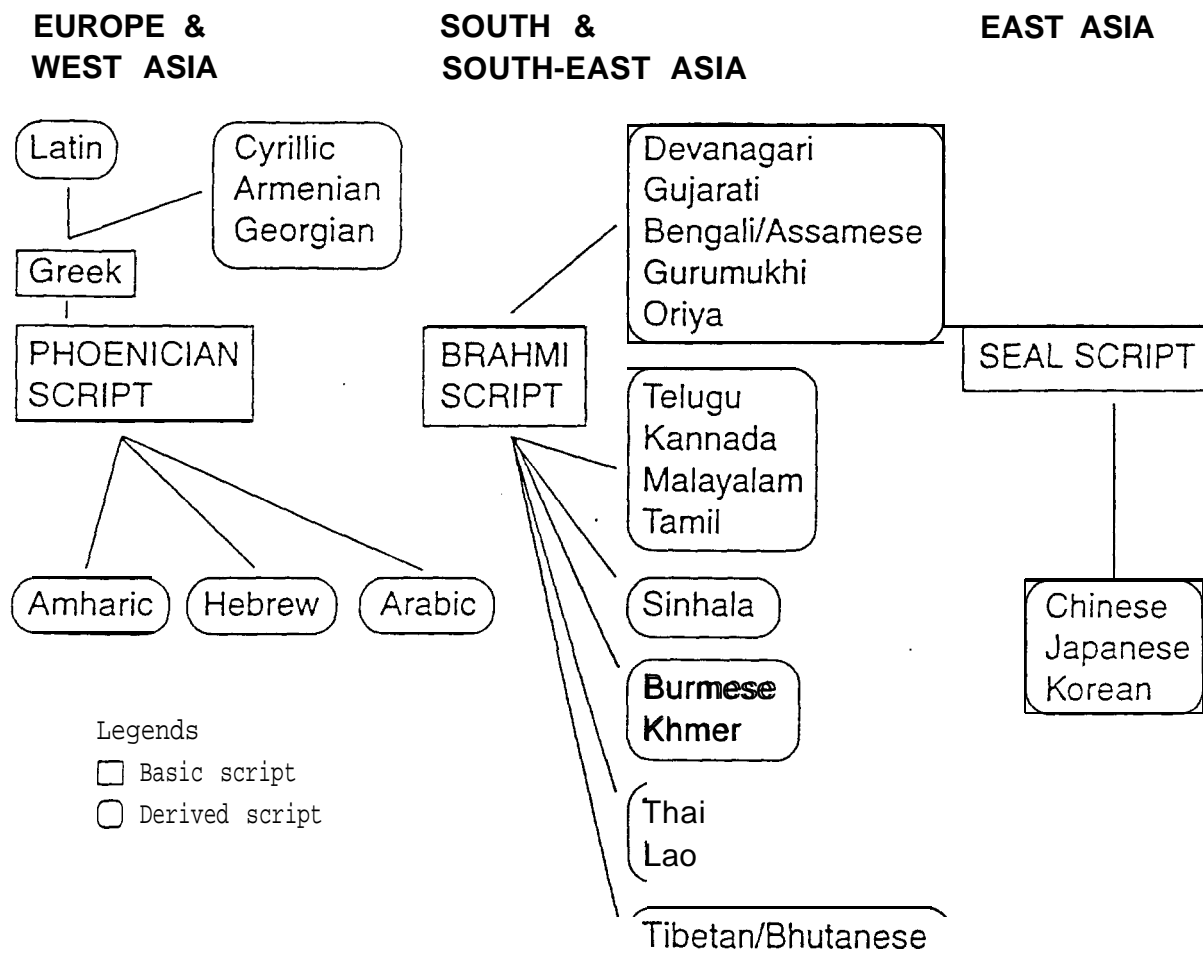
1.1 East Asia - Chinese influences

The oldest scripts are ideographic (symbols directly representing meaning rather than sound), and as such are still used in China. Chinese characters were adopted and adapted - with additional or alternative phonetic characters - in Japan and Korea. and at one time. in Vietnam, particularly when various religions such as Taoism. Confucianism and Buddhism spread from or through China at various times in the past.

East Asia has many great treasures of writings contained in manuscripts, but obviously copying of manuscripts by hand could never have led to the availability of publication we are used to today. Xylographs (wood-block printing) were developed, primarily to spread religious texts, many of the earliest coming from East Asia.

The East Asian written languages which use Chinese characters today - Chinese, Japanese, and to a lesser extent, Korean, draw on a writing tradition going back over millennia. The most recent changes to the written language

2 Figure 1: The three families of scripts in current use worldwide



3 Figure 2: Examples of Chinese characters. showing common radical features

瞄	睽	睿	睡	磁	碟	碧	碳	碩	碣
4001	4002	4003	4004	4005	4006	4007	4008	4009	4010
禎	福	禍	種	稱	窪	窩	竭	端	管
4011	4012	4013	4014	4015	4016	4017	4018	4019	4020
箕	筴	筵	算	箝	箔	箏	箬	箇	筭
4021	4022	4023	4024	4025	4026	4027	4028	4029	4030

reflect government backed standardization on the range of characters used at various educational levels. and in the People's Republic of China. a major simplification of the written forms of characters. Traditional Chinese characters are shown in Figure 2: this also shows an important way of indexing characters. The common radical element seen on the left of characters 4001. 4002 and 4004 also appears below the rest of the character in 4003 - a similar feature can be seen in the group 4005-4010. The common element can also be seen above the rest of the character in 4020-4030.

1.2 West Asia and Europe - Phoenician influences

At the other end of Asia, West Asia also provided the earliest script to be used in the Mediterranean region. The early Phoenicians, who lived in what is now Lebanon. were major traders. Their 22-letter script influenced nearby scripts like Hebrew, as well as later influencing European scripts like Latin. Greek. Cyrillic, Georgian and Armenian. However. some features of this earlier script can also be seen in Arabic script. which has been used not just in the Arabian peninsula, but has also spread to other parts of West Asia, such as in Iran, and the Perso-Arabic tradition and the Perso-Arabic script had a major influence in parts of South Asia and Central Asia too.

Because the original Phoenician script was so influential that it was the basis of European scripts. which were the first to develop printing using movable type, these influences on written language are further explored in the rest of this section.

1.3 Europe and America - Christian influences

Just as Phoenician script, and other contemporary variants were adopted for Hebrew script. Hebrew in turn was adopted and adapted for writing Greek. which again in turn was adopted as a model for writing the Latin, Cyrillic, Georgian and Armenian scripts, each of which spread under the influence of Christianity. Biblical texts and commentaries indeed formed a major part of early printed publications, and indeed this provided a major stimulus towards the early development of printing in a wide range of scripts. Much of the earliest non-roman script printing was in fact undertaken by, or at the behest of. the Vatican in Rome for publishing and distributing bibles and other religious texts overseas: examples can still be found in the Vatican's libraries and archives for many Asian languages.

Latin script languages, were the first to develop movable-type printing, which led to a publication revolution worldwide. Countries using other scripts with similar characteristics to Latin script - using separate rather than joined-up characters - such as Greek and Cyrillic - were the first to benefit from this technology, but typographers soon turned their attention to most other scripts in use worldwide. By the time that printing was mechanised in the nineteenth century. with large numbers of trained machine operators controlling hot metal processes to produce regular publications in large quantities, all current languages have had typefaces developed for them which were used both in the West and in Asia.

1.4 West Asia, South Asia and Central Asia - Islamic influences

Arabic can also be said to have some resemblances to the Phoenician script, although Arabic uses several more letters in its alphabet, and also a word-root system, predominantly using three-letter roots, as well as providing an option to show vowel signs above or below letters.

These vowel signs are used primarily in religious and educational texts, where total unambiguity of meaning is vital. "In general publishing however -

as in newspapers - the lack of vowel signs does not give rise to any problems because several letters also tend to take on certain vowel functions and in any case the larger the amount of words. the more the context tends to resolve any ambiguities.

Arabic is written from right-to-left, and with letters mostly joined together within words. and with a range of ligatures. so that letters had a variety of different shapes depending on their position within the word. When printing was developed for Arabic, some of these character shapes and ligatures were simplified and standardised in order to avoid printing becoming a very slow process, and to allow manufacturers to supply faster machinery to cope with the growing publication demands.

The spread of Islam outside Arabic-speaking countries led to additional characters being added to the Arabic alphabet. for instance in Farsi and Urdu. An example of Arabic script extended to other languages can be seen in Figure 3: here the additional letters adopted for Farsi have been further added to for Urdu. This figure also shows the different forms letters can take within words - a feature that applies to all Arabic script languages.

Various Turkic languages (e.g. Kazakh, Kirgiz, Uighur) used in Central Asia and parts of China also adopted Arabic script. As with the Arabic language, most publications do not use additional vowel signs in these extended Arabic scripts. However, since the Kirgiz, Kazakh and Uighur languages in China adopted Arabic script, certain letter forms have been used consistently to represent vowels in these languages.

1.5 South and Southeast Asia - Hindu and Buddhist influences

Most South Asian and Southeast Asian scripts represent vowel sounds by vowel-sires above or below letters, as in Arabic, but in this case their use is mandatory. They are all written from left-to-right. All South Asian alphabets follow a very logical phonetic order - so logical that this was the order adopted (with modifications) by the International Phonetic Association (IPA) for the International Phonetic Alphabet.

Several of these scripts also combine letters in quite complex ways - certainly compared to Latin-based scripts where letters in words mostly follow each other without alteration. Figure 4 shows how basic letters (above) are transformed to a more complex word form (below) in Devanagari script - that used for Hindi, Marathi and some other North Indian languages. All computer equipment for Indian languages needs to provide for this degree of complexity in both display and printing, without adding any extra complexity to the keyboard or other input system.

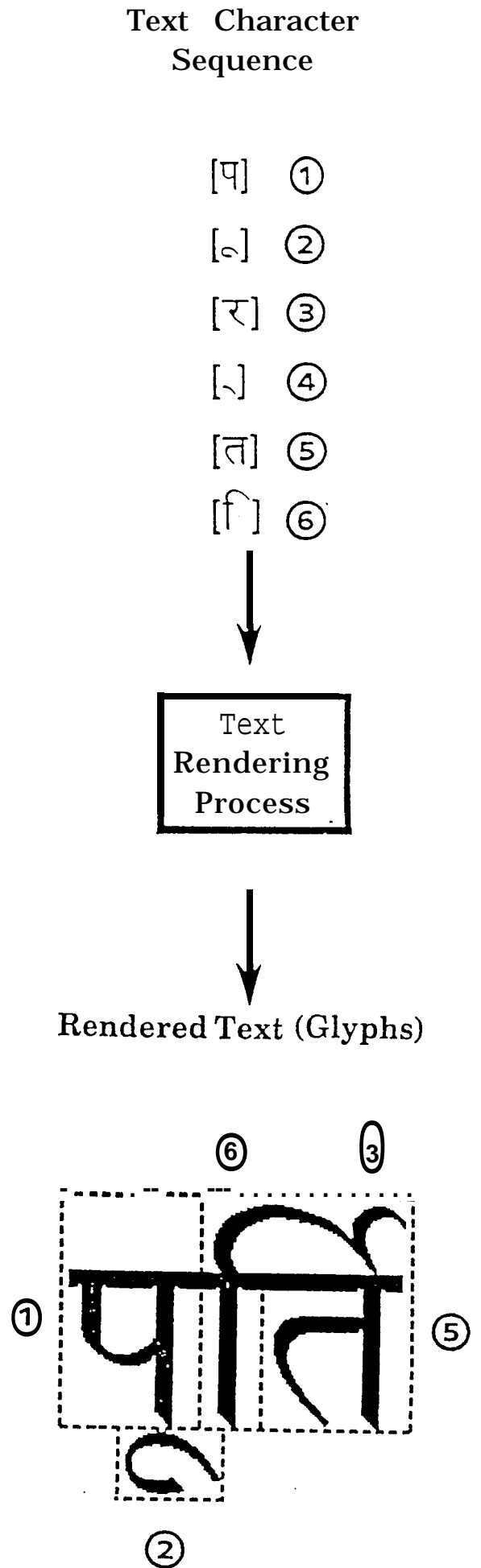
In South and Southeast Asia. as elsewhere in the world, scripts have tended to follow religious and other cultural influences. The spread of Hinduism, and later Buddhism, meant that the Brahmi script (and later Devanagari script, as used in Sanskrit and Hindi) had a major influence on the scripts used in other parts of South Asia and Southeast Asia. and a similar alphabetic order and other similarities can easily be traced in all the non-Latin scripts used in India, Bangladesh, Sri Lanka, Tibet, Burma, Cambodia, Thailand and Laos. Burmese, Khmer, Thai and Lao also use tone signs to disambiguate words, reflecting the use of tones in speech for the same purpose.

In some of these languages, some letters are no longer used, and this is reflected in the computer character set standards, where they exist, which do not usually code such letters.

Figure 3: Arabic letters in Urdu: lone. initial. medial and final forms

ا	ب	پ	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ک	گ	ل	م	ن	و	ہ	ی	ے
ا	ب	پ	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ک	گ	ل	م	ن	و	ہ	ی	ے
ا	ب	پ	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ک	گ	ل	م	ن	و	ہ	ی	ے
ا	ب	پ	ت	ث	ج	ح	خ	د	ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ	ف	ق	ک	گ	ل	م	ن	و	ہ	ی	ے

Figure 4: How alphabetic data relates to graphical information in Devanagari



2. Language planning in modern times

Languages have continued to be reformed over time. particularly in areas of spelling and pronunciation, where governments have had a major role to play through educational and other planning processes. Printing and publishing themselves have also had a defining role in the use of language. and in some cases computer technology has had a further defining role. in both cases affecting the range of characters used.

2.1 East Asia

In China. the reform of some of the more complex traditional ideographic characters to simplified shapes is perhaps the best known example of language planning in East Asia, along with the use of pinyin in romanization. This latter was envisaged by some as a potential alternative to Chinese characters to increase literacy. but the large number of homophones (words with different meanings but similarly pronounced) means that it will not replace Chinese ideographic characters for most purposes. China also has made major changes in some of the other languages used in China, mainly those used in its Southern and Western provinces. These have meant providing new orthographies (writing systems) - mostly based on the Latin alphabet. as in Zhuang, or in resurrecting, adapting and standardizing existing orthographies. as in the Yi syllabary, the traditional Mongolian alphabet, and the Arabic alphabet adapted for Kazakh, Kirgiz and Uighur.

Among the countries and territories that use Chinese characters (hanzi) in their script, China is not alone in language planning. Hong Kong and Singapore have both adopted the pinyin transliteration and the Mandarin spoken dialect for official use, although Taiwan's recent changes to the official transliteration system have been made independently of other Chinese speaking countries. Only the People's Republic of China has adopted simplified characters: in Hong Kong people's familiarity with traditional characters means that some attention needs to be given to this matter. Several people in the Chinese Academy of Sciences who are concerned with language planning have stated over the years that some of the characters may have been oversimplified: whether the re-incorporation of Hong Kong into China in 1997 means that this may be an occasion to re-rationalise the use of simplified and traditional characters within China remains to be seen. Fortunately. many publishing computerised systems on the market enable printing in both simplified and traditional characters.

Japan's use of Chinese characters (kanji) has gone through several reforms, mainly in specifying the range of characters that should be known by adults, and during various stages of the educational process. These largely match the first of two levels of characters available in the national character set standard. The standard transliterations - the official kunrei and the widely used Hepburn (Hebon-shiki) transliteration systems, as well as the phonetic Japanese kana syllabary are fortunately so standardised, and so unambiguous that all three can be use as input systems in many Japanese computers and publishing systems.

In Korea, use of Chinese characters has always been much lower, given that there are fewer homophones than in Chinese, and that the Hangul script is completely phonetic. In the Republic of Korea, there have been movements to eliminate the use of Chinese characters (hanja), although most publications use a certain amount. In the Democratic People's Republic of Korea, only the phonetic Hangul characters are used.

Figure 5: Extended Devanagari repertoire approved in 1989

Vowels :

अ आ इ ई उ ऊ ऋ लृ
ए ऐ ओ औ अं अः

Matras :

ा ि ि ु ू े ै ो ौ ः

Consonants:

क ख ग घ ङ
च छ ज झ ञ
ट ठ ड ढ ण
त थ द ध न
प फ ब भ म
य र ल व
श ष स ह ङ ढ ळ
क्ष त्र श्र

Figures :

१, २, ३, ४, ५, ६, ७, ८, ९, ०

Explanations : Conjunct letters :

(1) Consonants with vertical line (बड़ी पाई) :

ख ग घ च ज झ ञ ण त थ ध न
प ब भ म य ल व श ष स क्ष ञ

Conjunct letters, in case of these consonants, should be formed by removing the vertical line e.g. स्वीकृत, यक्ष्मा ।

(2) Other consonants :

- The present form of the conjunct क and फ should continue, e. g. संयुक्त, पक्का, दफ़तर
(not as संयुक्त, पक्का, दफ़तर)
- The conjunct forms of ड छ ट ठ ड and द should be made by adding the ह् symbol (ह्), e. g. वाङ्मय लूट बुद्धा विद्या etc
(not as वाङ्मय लूट बुद्धा विद्या)
- All the three old forms of conjunct र should continue; as प्रकार, धर्म, राष्ट्र
- श्र should be written in the old style as in श्री.
- Conjunct form of त्र and र should be written as त्र instead of त्र (Eater, the form त्र as in vogue earlier, has also been approved).
- Conjunct consonant with ह् . may also be formed with ह् symbol besides that in vogue e.g. विद्म and चिह्न (but not चिह्न).

2.2 South Asia

India's nine official scripts are largely based on the Brahmi script, and closely related in their repertoire and alphabetic order to Devanagari script, as used in Hindi and several other North Indian languages. However, the dominance of English and the reluctance in parts of India to adopt Hindi as an official language - to many as foreign a language as English - meant that English has had a major part to play in Indian publishing, with a wide English speaking market for publishers.

5
6
7 However, the English language heritage also had other effects on Indian language planning. In some cases substituting Latin script was seen as a solution to India's multilingual situation: during the late 1920s, during colonial times, some linguists proposed an extended-Latin All-India script, based on similar moves in British West Africa, and with similar letters. In the early 1980s the Roman Lipi movement used the relative unavailability of Indian script printing and computing capability as an argument for promoting Latin script transcription of Indian languages for general use.

Neither of these made any impact, and indeed the Government of India has had a major role in language planning the written language and its use in computers - one that might serve as a model for other governments.

8.9 A standard Devanagari script - avoiding the complex compound characters - was defined at the 1959 Education Ministers' Conference, and accepted by the Government of India. In 1966, the Government of India's Central Hindi Directorate produced an extended - Parivardhit - Devanagari to allow for other Indian languages to be represented in Devanagari (an approach which in principle would also allow all Indian languages to be represented in any of the other Indian scripts, such as Gujerati, Gurumukhi, Bengali, Oriya, Tamil, Telugu, Kannada and Malayalam). This approach of mapping one script to another was also taken by the Bureau of Indian Standards, in its development of the ISCII standard, and by the Government of India's Centre for the Development of Advanced Computing (CDAC) in the development of its GIST hardware and software widely used throughout India for data processing, word processing, telecommunications, publishing and in the film industry.

2.3 West Asia

The Arabic language and script has remained very standardised in its use, through its association with Islam, rather than through government planning. In printing, Arabic has used several forms for each letter to allow for the variations in shape which flow from handwritten Arabic - much as handwritten English letters can alter if they begin rather than end the word. Figure 6 shows how letters should appear following Arabic calligraphic conventions.

10 Around twenty years ago, a few typographers attempted to fit Arabic script into the straight jacket of European language printing systems, which gained some intellectual credibility, but much popular opposition. At the time, and these proposals are now largely abandoned. Some examples of these are shown in Figure 7. All these are simplified: the letter H (beginning each line at the right) should have different form within the word; the letter N (letter with a single dot above in all examples) shows more adaptation to Arabic norms in all but the last example, where it appears (whether in the middle or the end of a word) like the letter i in the Latin alphabet. As with arguments over Roman Lipi in India (see above), computer usage of Arabic script when it developed took on board much of the printing conventions of different forms of letters, and indeed computer programs now exist to allow computers to emulate the most demanding of manuscript styles, so that printing using the repertoire of letter forms as in figure 6 can be expected from most systems.

Figure 6: Arabic calligraphic conventions in a computerized system

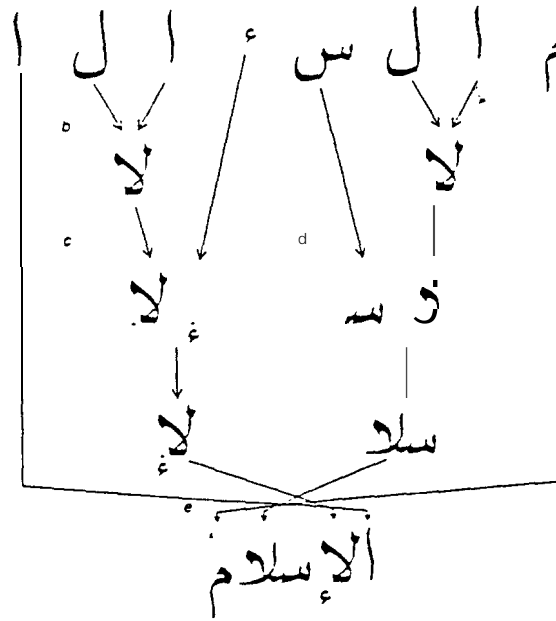


Figure 7: Forcing Arabic script to follow Western printing. 1950s - 1970s.

هـنـيـهـة مـنـ أـلـتـفـكـير
هـنـيـهـت مـنـ التـفـكـير
هـنـيـهـة مـنـ التـفـكـير
هـنـيـهـة مـنـ النـفـكـير
هـنـيـهـة مـنـ أـلـتـفـكـير
هـنـيـهـة مـنـ أـلـتـفـكـير

2.4 Southeast Asia and Oceania

In the nineteenth century colonial powers imposed Latin script conventions on languages in several countries they occupied. This was particularly true in Africa. where Latin script mainly followed English or French models of Pronunciation and writing. In Asia this approach was also applied to Vietnamese, although the European letters and diacritics (borrowed from French and Portuguese models) did at least fit the structure of the Vietnamese language, and the same system has remained in use since independence.

Written languages in the Philippines are also Latin based. but with a much simpler alphabet. due to Spanish and American influences in colonial times and the alphabet of Filipino (based on Tagalog) and other languages of the Philippines have avoided most accents. and indeed their phonetics are much simpler than Vietnamese.

In Malaysia and Indonesia. although languages such as Javanese have been written in scripts derived from Brahmi (see above) these are no longer currently used. Because of the predominance of Islam in both of these countries. Jawi script (Malay written using Arabic script) is still used in some publications. However, rather than adopting Arabic script for official use, both countries together adopted a unified simple English language script for their languages (known respectively as Bahasa Malaysia and Bahasa Indonesia) during the 1950s which meant that adoption of standard Western printing technology - and indeed computer technology in general - has been very straightforward.

In Oceania. their proximity to other Latin script countries, as well as post-colonial trade links, led the island communities of the Pacific down a similar route: where their languages were written it had always been in Latin script, and strong economic links with Australia. New Zealand and the USA meant that there was no reason to change this.

2.5 Central Asia

The one area where some changes may still occur over the next decade or so is Central Asia. Islamic influence meant that the predominantly Turkic-language peoples in this area traditionally used Arabic script. However, various secular influences changed the use of scripts. In the 1920s. several primarily Moslem republics adopted Latin script in the cause of modernization, in particular Turkey and some republics in Soviet Central Asia - indeed a Unified Turkic Latin Alphabet was used.

Under pressure from Moscow, the Soviet Central Asian republics adopted Cyrillic script with a bewildering variety of additional Cyrillic letters, as indeed did the Mongolian People's Republic. Conversely in China. in response to public pressure, Turkic peoples in Xinjiang province moved from Latin to Arabic script.

Following the break-up of the Soviet Union, Mongolia has chosen to abandon Cyrillic script in favour of the traditional Mongolian script, as do the Mongolians in Chinese Inner Mongolia, where the vertical Mongolian script has always been used.

Azerbaijan, adjacent to Turkey is the only independent Central Asian republic to have returned to the use of Latin script: in other Central Asian republics like Kazakhstan. Cyrillic script is still used for languages like Kazakh, and this seems likely to continue for the foreseeable future. One major influence here is the continuing presence of a large number (up to around 50%) of primarily Russian speakers in these republics. These republics

have come to an agreement with Turkey on what repertoire of Latin characters they will draw on if and when they do revert to Latin script - no calendar date has been set for this to occur. although a conference in Ankara in November 1992 set the scene for future developments. Turkey has been pressing for ISO character set standards to make this complete character repertoire available in the largest number of character set standards.

3. Technological influences on publishing in Asia and elsewhere

Technology was necessary for books, and later periodicals, to make any major impression - the copying of books by scribes prior to the invention of printing could make little impact beyond monasteries. Printing developed in Asia as elsewhere, in three main stages.

3.1 The first publishing revolution - movable type

Although wood-block printing was used in several parts of the world, it was the development of printing of books using moveable type, which allowed the publishing revolution to develop in the sixteenth century. Countries using Latin script languages were the first to pursue this revolution, which led to increased standardization in use of the script. This standardization approach was also seen when printing was adopted for other languages and scripts. and the printing revolution has largely been driven from the West, with Asian countries following Western models and conventions, and using Western equipment. Moveable type had enabled a major leap forward. and caused many more books to be written, and the first periodicals - mainly annual. or infrequent - appeared during the eighteenth century.

3,2 The second publishing revolution - industrialization

It was not until the nineteenth century, when the mechanisation enabled by the industrial revolution enabled the forces of steam, and later electricity, to be applied to the printing process and larger business enterprises were formed, that regular weekly or daily newspaper production could become widespread. This increased speed, and the colonial webs of commerce, led to the export to Asia of considerable numbers of mechanical printing presses and the growth in countries like India of a well-established publishing and printing industry.

However. although nineteenth-century mechanisation was still the same basic process but achieved much faster, it did allow new types of publishing. like newspapers, to emerge on a large scale during the nineteenth century. During the twentieth century the use of computing and telecommunications, and later computing power, did allow certain background processes to be much improved, as with news agency wire services and picture services. However, the end result is still printed publications.

3.3 The third publishing revolution - computerization

The computer revolution provides a new printing revolution, not just in enabling even faster/higher quality production processes, but having the capability for new ways of publishing data - e.g. in machine readable form, which can also be made available across data networks accessible by end users from anywhere in the world from the wide range of computers now in use worldwide. How this might develop during the next century can only barely be glimpsed, but publishing enterprises will need to keep a watching eye on continuing innovation, so that they and their customers can benefit from it.

Computerised publishing systems can now provide all the advantages of

document handling and data management required for large publishing operations, such as word processing, pre-press operations, incorporation of wire services from news agencies (for both text and pictures), archiving and retrieval. as well as a host of management features.

The conventions developed for moveable type printing have largely been followed when adapting to, and taking advantage of, the even greater opportunities for publishing offered by the use of computer systems in printing.

11 The power of computers and developments in the use of computer graphics in text handling mean that there is no longer any need to compromise the beauty of Asian scripts - providing the appropriate rules are built into the system design, any complexities can be left to the computer system, which can thus handle the same text in horizontal or vertical orientation (e.g. for Chinese or Mongolian), mixing of left-to-right text (as in European and South Asian languages) with right-to-left text (as in Arabic, Farsi and Urdu). Unfortunately, not all systems currently provide all language capabilities. However, publishers in Asia should be in a position to expect its provision when they are looking for new systems.

Part 2 - Key Language and technology issues

Part 2 of this background paper (sections 4-10) looks at how computers have been used in the publishing industry within Asia, and how well - or not - suppliers have addressed Asian language needs. It also deals with a number of technological issues which are likely to affect publishers in the future, whether they are faced with starting operations from scratch, or are faced with the need to update obsolescent equipment in the future. These issues include keyboards, character coding issues, telecommunications, and other issues affecting publishing, such as machine translation and machine transliteration.

12 Several of these issues, as they affect the general and research use of computers as these affect Asian languages are provided in the papers of the CPAL conferences supported by UNESCO.

Given the speed at which technology changes, and the reduced price of some computer technology, and the advantages which this new equipment provides, few companies can afford to ignore its impact and stay too long with older equipment. It also looks beyond the basic issues towards ways in which cooperation can help ad~ante publishing in Asia.

4. Computers in the publishing industry

13 In the 1960s and 1970s few publishers were involved in automation. Those that were initially used them for general administrative purposes, such as payroll: it was much later that the pressure, and the ability, to use computers in pre-press activities emerged. Rather than developing large mainframes or minicomputers, as in other industries, newspaper publishers were more inclined to invest in proprietary systems geared up just for newspaper production, by large companies with a track record in the publishing industry, such as those developed by Linotype and Monotype. For Asian users these companies were geared up to providing Asian language solutions? unlike mainstream computer companies.

4.1 Mainframes and minicomputers versus proprietary systems

Companies like Linotype and Monotype also provided a lot of technical support, and initially much of their computerised solutions also supported partially mechanical or chemical processes such as phototypesetting. These systems could not be used directly with other computerised systems that the business might have, for general administrative purposes etc. The general business market by comparison was dominated by multinationals such as IBM, Digital and Hewlett-Packard. However, their mainframe and minicomputer business systems made less impact on the newspaper industry than in other commercial areas.

Even the UNIX operating system, developed and supported by several vendors to allow for transportability among minicomputers, did not make an enormous impact here. Although several UNIX systems were adapted to make use of Asian language data, generally these were in specialist departments of large universities, and they had a minimal impact on Asian business needs. Despite any limitations in microcomputers, considering the vast increases in computing power and the rapid decreases in cost, "downsizing" is now the buzzword for many organizations - even larger ones. Some of these - even university computer departments - are completely getting rid of their mainframe computers in favour of PC networks. In some application areas, developers have even wondered whether it is worth their while developing UNIX versions of their software, considering the increasing power of microcomputers.

14 4.2 Microcomputers - the Macintosh

In the late 1980s and 1990s, microcomputers became commonplace, one per desk in many businesses. With the possibility of linking several together in a network, as larger and more expensive terminals had been linked to mainframes and minicomputers had been used previously, and with the large number of different applications that could be run on them, and with the continuing drop in cost due to increasing sales and increased competition, the microcomputer became an increasingly useful tool for all businesses.

Newspapers were among the first to make use of this, particularly with the Apple Macintosh (or Mac). The development of much higher quality printing associated with Apple technology enabled it to play a leading role in newspaper production. Several newspaper equipment suppliers like Linotype and Monotype used the Mac as a front end to their proprietary system, particularly as its Graphic User Interface (GUI) meant that it was easier for new users to pick up than the text-based, user-impenetrable mainframes, minicomputers and the microcomputers. Today, even the Mac alone, with appropriate high quality printing, provides sufficient high quality desktop publishing (DTP) facilities for many newspapers, through continued improvements to the Mac. and add-ons like Quark Express, and Adobe Acrobat etc. often in association with other hardware and software (see the section on Systems Integration below).

The Mac has always had good language processing at the heart of the system, through its Script Manager. which allows most languages and scripts to be used, although not necessarily simultaneously, and there are long-established suppliers of Asian fonts for the Macintosh. With the development of the Macintosh System 7, together with Quickdraw GX, full UNICODE support is also available, offering many Asian scripts simultaneously, which is of particular use to publishers of multinational and multilingual publications.

Furthermore, there is a wide range of other applications, such as word processing, database and spreadsheet applications that can be of use to a newspaper business and can in principal be integrated with other application software to integrate many business functions within a newspaper.

15 4.3 Microcomputers - the PC

The IBM PC is almost a misnomer nowadays, in that due to more open licensing arrangements, other manufacturers were able to produce similar PC "clones" to the IBM PC, and there are more non-IBM PCs being used than IBM-badged PCs. The generic PC has continued to be developed, with other chip manufacturers (such as Intel) and software developers (such as Microsoft) dominating hardware and software development respectively.

16 Only five years ago, the Mac had a clear lead in most of these areas which affect the printing and publishing industry. The only lead PCs had were in the wide range of software written for them - much of it not usable or integrated in conjunction with other software.

However, PCs now equal - and arguably lead - Mats in most of the areas listed above. The Microsoft Windows GUI provides a similar operating environments to the Macintosh GUI, and the associated programming environment has led to a wide range of software, interoperable within the same computer. The widespread use of networking in computers has also led many third-party software developers to produce client-server versions of their mainframe software, which enables large companies to make much more use of cheaper microcomputer power within large-scale business operations than previously.

For Asian scripts too, there are now a wide variety of high quality fonts available. Both for newspaper companies who are intending to automate now, and for newspaper companies with Macintosh-based hardware and software still performing to requirements who may need to upgrade in the future. It is worth keeping an lookout for future trends. The Macintosh is entrenched in current systems, and newspaper specific software is still being developed by companies such as Diwan, but these are also developing parallel Windows-based versions of their software.

4.4 Systems Integration

Systems Integration is the "bolting together" of systems which were not themselves designed to work together - even PCs and Mats linked together. While for many years Open Systems have been pursued as a major objective, so that in principle any types of hardware and software can run together, the reality is that Open Systems still have specific limits as to which components will work together - Open Systems are only partially open. However, current technology and the current marketplace has led to a series of third-party alliances and cooperative ventures so that with an experienced system integration company can provide a computerised solution to provide for the needs of any enterprise.

Typical systems integration in the Newspaper industry might be the use of either IBM PCs or Apple Macintosh computers (used as "front ends" to a powerful SQL database system mounted on a mainframe or minicomputer system, incorporating internal word processing, receipt of external data via wire services, advertisement preparation, incorporation of pictures with text, page make-up and all other pre-press activities. transmission of data to the printer, either locally or remotely, via cable or satellite links in order to speed distribution! archiving provision and management functions. which could be from several suppliers.

17 Examples of larger systems integration in West Asia include a number of Arabic papers where Diwan (based in London) has provided the systems integration, where the Apple Macintosh and/or Microsoft Arabic Windows on PCs have been used as the front end. although Apple, UNIX or DECvax servers can be used for some of Diwan's current NewsNet, CommsNet and PictureNet applications. Integrating pictures with text is a major concern in today's newspaper market, as is multi-site printing, using telecommunications (see also section 7 below).

18 An example of a smaller and lower cost system, with some of the above functions, with single-site (local) printing is that of the Dainik Free Press which is a daily Hindi edition of an existing English language newspaper> launched on January 17, 1994 by the Free Press Group with a starting daily circulation of 20,000. The Dainik Free Press uses CDAC's GIST technology to allow reporters and other staff to use 15 of its basic low-cost DOS word-processors for data input, as well as using external news-agency information in Hindi, which uses the GIST card to process Hindi data. At a later stage of production, information from these sources is fed into 6 computers using CDAC's higher level DTP and page make-up software under Windows. This is then merged at a final stage in the daily schedule at a single terminal on the chief editor's desk (Figure 8). A smaller scale weekly paper for Tibetans living in exile in India had also been set up, but had not reached the critical mass necessary for uninterrupted production.

Manufacturers of proprietary systems such as Linotype might argue that their systems have also provided systems integration for some years, albeit largely using their own proprietary hardware, software, interfaces and conventions (see the composite typical configuration in Figure 9). There are still large

Figure 8: Systems integration in a PC-based newspaper system (by CDAC, Pune)

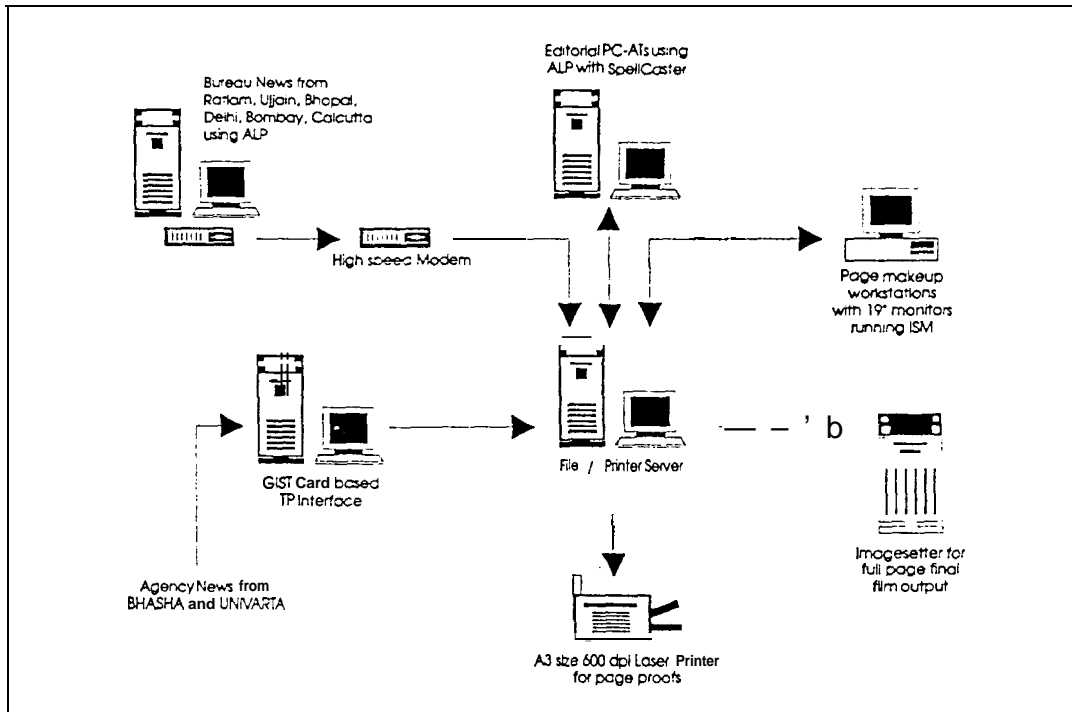
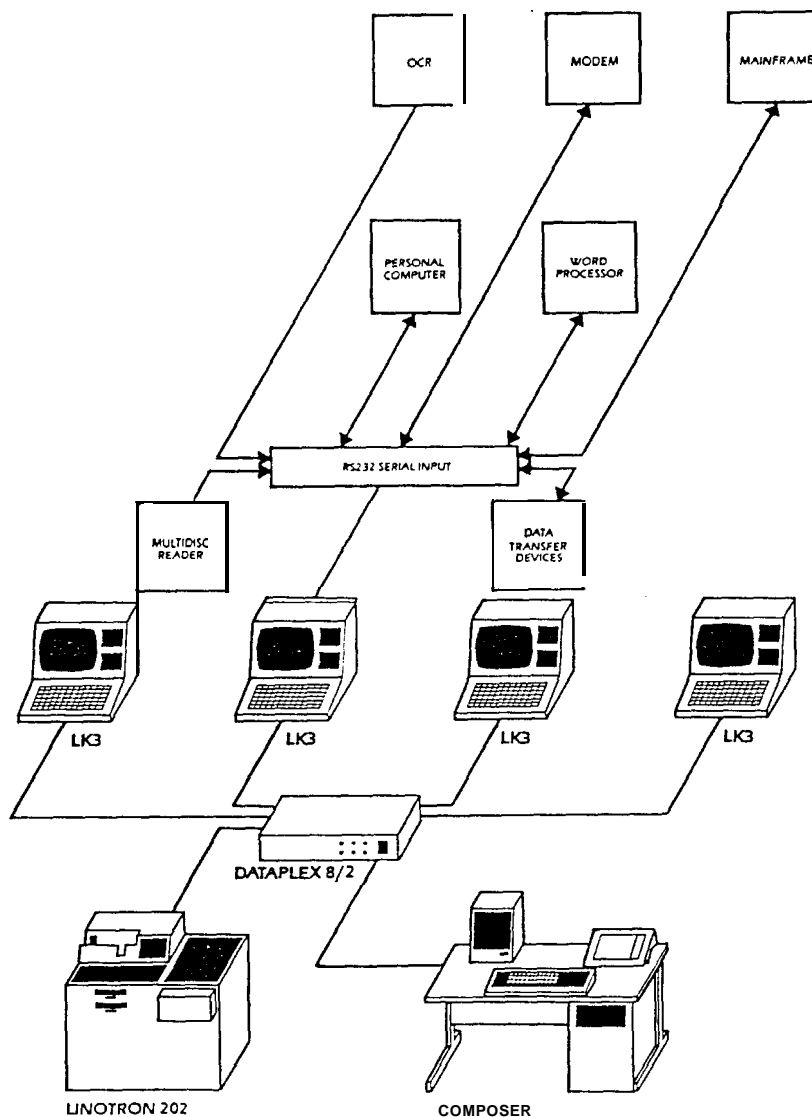


Figure 9: Systems integration in a Linotype newspaper system



numbers of these older "legacy" systems in Asia, which have provided large-scale publishing systems in several Asian cities covering most of Asia's languages used at national or federal state level.

4.5 Future concerns: technological issues which affect publishers

IBM PCs and Apple Macintoshes have both dominated the microcomputer Market for the last ten years, with several manufacturers dominating the mainframe and minicomputer market. particularly IBM. Hewlett-Packard and Digital, as well as companies like Bull and Unisys. Hewlett-Packard and Epson have also dominated the computer printer market. The suppliers of large newspaper printing equipment have also relied on using some of these, particularly with Macintoshes as front-end printing systems.

After a decade or so of stability dominated by the PC and Macintosh environments. 1995 also represents something of a crossroads in computer development. Not only have the Apple Mac and the PC been considerably expanded from their original architecture, so that software designed for the latest Mac or PC cannot be guaranteed to run on the earliest Mats or PCs, but their are new. more powerful types of computers now being produced which may enable completely new types of software to be developed. Examples are the DEC Alpha computer, and the PowerPC chip, jointly being developed by IBM. Apple and Motorola, which have already resulted in a Macintosh PowerPC computer, with IBM likely to release an IBM PowerPC computer soon.

New operating systems are also being released: as well as Apple's Macintosh System 7, and Microsoft's Windows NT - both of which can use UNICODE encoding - Microsoft is also releasing Windows 95, and IBM releasing its WARP operating system, which is a major enhancement of its OS/2 software. Continuing upgrades can also be expected for proprietary operating systems {mainly in Mainframes and minicomputers}.

The fifteen years since the advent of the microcomputer is a long time for a computer design - the IBM PC was only expected to last a year or so before being replaced by its proposed successor, and old computer architecture places limitations on future developments. New chips (microprocessors). computers and operating systems are already being developed, for instance the DEC Alpha RISC computer from Digital, and the PowerPC, jointly developed by IBM, Apple and Motorola. In addition there are new operating systems like Windows NT, Windows 95, IBM's WARP (the latest OS/2), as well as others like NeXTStep and Solaris. Object Oriented Programming is also likely to have a major impact on the capability of systems to handle a wide variety of languages and scripts. All these developments mean that systems are likely to change significantly over the next decade, and purchasing decisions made now need to be made "future-proof" against any developments which will make current systems increasingly obsolescent.

As this affects newspaper publishers, it may well be useful to seek outside assistance in looking at how these trends may have developed when it comes to purchasing new hardware or software for newspaper publishing, to ensure that whatever is purchased is likely to continue receiving support.

The above may discourage newspaper publishers from investing in computer technology at all, but the benefits that can be obtained from using computers in the newspaper business, as in all businesses, mean that those who do not invest in computers will in due course be overtaken by their competitors who do.

Specific issues which Asian publishers should consider when acquiring new equipment are covered below. The principal language related areas are

keyboard layouts to ensure efficiency of staff; character coding to ensure ease of information transfer between and within systems, and also related telecommunications issues: sorting and retrieval issues for indexes and directories in publications, and to ensure full access to newspaper archives; and other issues which may arise in the future, such as translation and transliteration issues.

5. Keyboards

Increasingly more and more people involved in publishing and printing are working at a keyboard for much of the day. There are several issues relating to keyboards which are important when acquiring equipment. as over a period of time they will affect the costs of the publishing operation.

5.1 Keyboard shape

19 Normal typing requires holding the wrists continuously in an unnatural position. In the West there are an increasing number of legal cases resulting from disabilities caused by Repetitive Strain Injury (RSI) associated with overuse of particular muscle groups, often diagnosed as conditions such as tenosynovitis, carpal tunnel syndrome, epicondylitis, etc. This is expensive to employers, whether through the costs of the absence, or slowing down, of the trained worker, and the need to replace them by other staff who will need retraining, or in extreme cases, through the costs of hospitalization and/or litigation, as in several recent cases in the West. Fortunately this problem can be eliminated before it arises if appropriate keyboards are used: unfortunately few manufacturers supply them as standard with computers or printing systems, but they are available as add-ons in most cases.

20 The Maltron keyboard has had the most impact in this area. Its main feature (Figure 10) is its concave design and angled keys which allow the wrists to lay more naturally, and avoid RSI. Originally designed with an unfamiliar arrangement of keys (similar in principle to the Dvorak keyboard in Figure 21 11) to allow for a more even load on the fingers than the conventional QWERTY layouts (also shown in Figure 11) its sales have increased since QWERTY layouts were also made available on a concave, angled keyboard.

A certain number of foreign language versions have been produced, but so far only European language versions (with accented characters etc) although in principle there is no reason why other existing keyboard layouts for Asian languages should not be superimposed on this concave model. The preferred Maltron Latin keyboard layout is shown in figure 12, and loading comparisons with Dvorak and QWERTY keyboards in figure 13 - a standard analysis that should be performed on any keyboard layout? for any language or script, to determine its efficiency. Long established use, as in QWERTY and its variants for Latin script keyboards, is also a major factor as is logical order and/or memorability, as in figures 14 and 15 (see also section 5.2)

Maltron also produce one-handed and other keyboards which can be used by disabled people. As office work is often one of the few openings for disabled people, as their disability can limit the amount of physical work they can do, and that some governments pass legislation regarding employment opportunities for disabled people, and that some companies in any case make special work opportunities available in offices, this approach may be of interest to newspaper publishers given that Maltron keyboards exist for PCs, and are also developed for a variety of specialist terminals.

Some of the most recent Apple Macintosh models now come with a keyboard that can be split down the centre from a hinge at the top, also allowing the hands

Figure 10: Maltron keyboard designed to avoid RSI

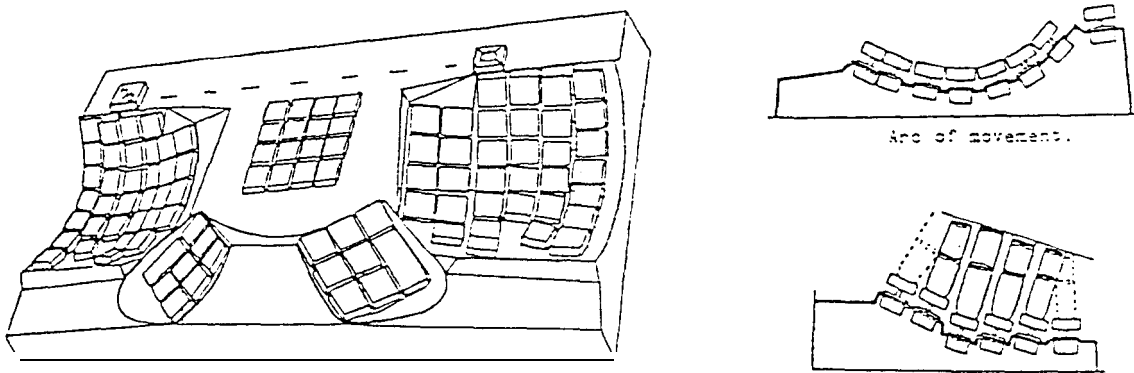


Figure 11: Keyboard layouts: comparison of Dvorak and QWERTY arrangements

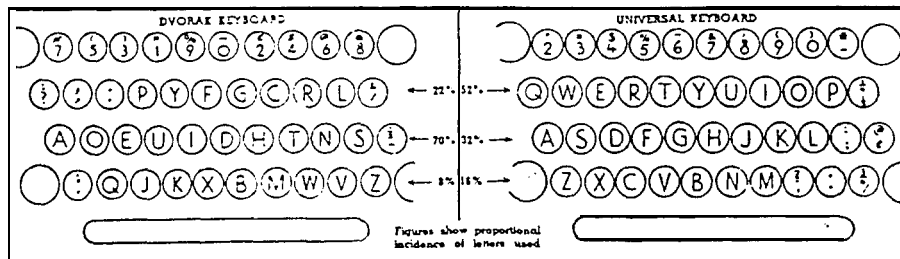


Figure 12: Recommended Maltron layout for optimum finger loading

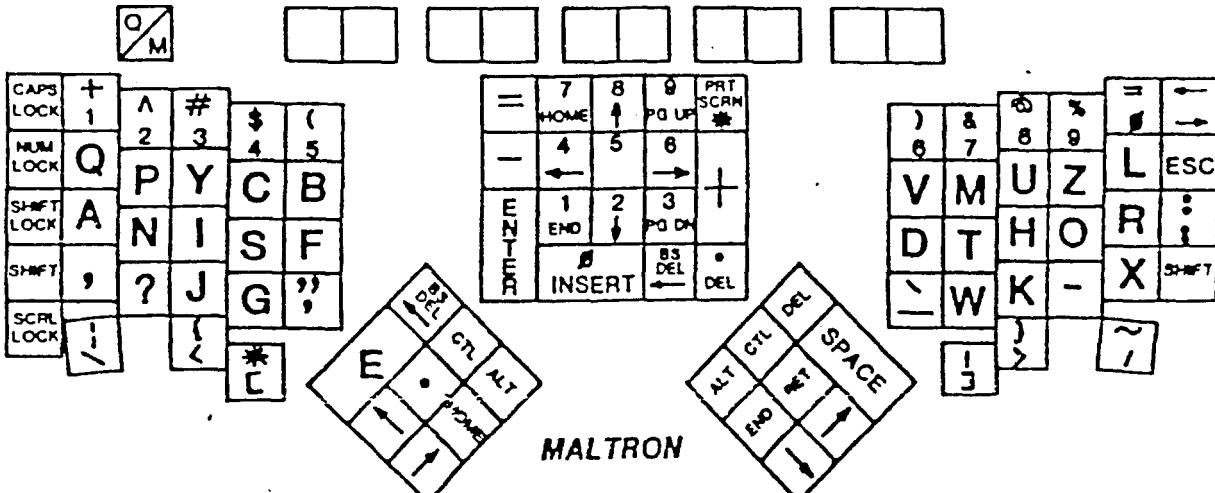
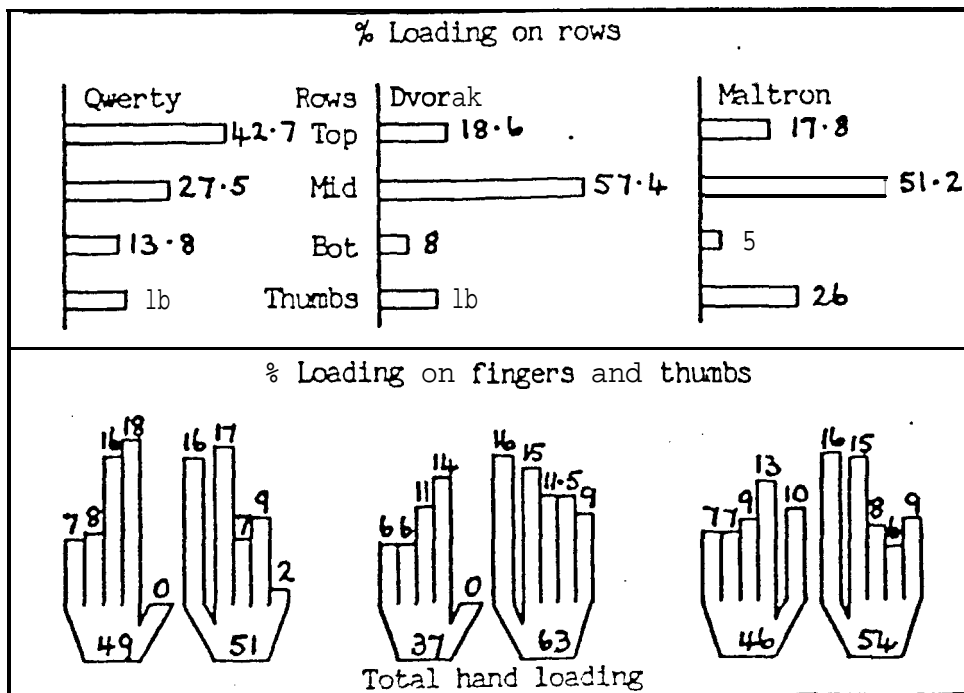


Figure 13: Comparative row loading and finger loading of the above keyboards



to work in a similar position to that of the Maltron keyboard, which is useful considering the predominance of Macintoshes in newspaper publishing, and the availability of non-reman script keyboards for Macintoshes.

5.2 National layouts ("typewriter layouts")

Even "two-fingered" typists familiar with a keyboard layout can work faster than typists not familiar with the keyboard layout. The cost of a number of slow typists in an organization tends to be a hidden cost which multiplies over a period of time, but which accountants seldom acknowledge.

22 Most scripts have national layouts (frequently called typewriter layouts)
23 which have been optimised to allow the most frequently used letters to be
24 most accessible on the home row, and/or to allow less frequently used letters
to be found in an obvious place. Figure 14 shows the Indian national standard
layout, with the keys logically arranged in vowel and varg groups: the
prevalent Arabic layout in Figure 15 has clotted letters adjacent to related
undotted letters, based on a long-standing similar arrangement taken from
Arabic typewriters. A standard layout is useful so that people do not need
retraining when they move to different equipment in the company, or to
different jobs.

Typewriter layouts are usually fine for computer keyboard layouts too. although the fact that computers can do some of the donkey work, such as automatically determining correct forms of letters - previously done by shift key(s) or through allocating more letter forms on spare keys - means that computer keyboard layouts can be simpler than their typewriter equivalents. Examples this are Arabic. where only one letter per key is now required, and Chinese, Japanese or Korean, where a standard computer keyboard can be used to input characters, compared to previous very large typewriters with a layout of hundreds of keys.

25 In China, Japan and Korea there are essentially four types of input systems: Phonetic input; Numerical input; Character component input; and Alphabetic coding. Some users prefer one type: others another. Large keyboards were the norm in these countries for early computers, following the large typewriter tradition - sometimes using multiple shifts, as in Figure 16. However. most East Asian language computer systems now have a choice of input methods. In Japan and Korea most input is phonetic, with software converting phonetic characters to Chinese characters if the user exercises that option, and some Chinese computers even allow touch typing of characters from pinyin input. Most input systems can now work from a standard PC keyboard.

5.3 Phonetic layouts ("QWERTY layouts")

For computer keyboard layouts, most Asian countries either have standards, or well established traditional layouts based on the dominance of one particular typewriter manufacturer in the past. In some countries (a) no particular traditional keyboard was predominant for either typewriters or computers, or (b) several different languages are used in that country, or (c) the use of English (or sometimes French or Portuguese) as a second language is very prominent. In the last two cases the use of a national keyboard layout as well as a QWERTY layout, or an AZERTY variant, means that computer users need to memorise two different layouts, frequently without one set of letters being engraved on the keyboard. In all three cases, there is an argument for considering whether a QWERTY equivalent layout for the Asian script letters is worth using: knowing one layout well that will serve a range of languages should be more effective than knowing two layouts less well.

Figure 14: Standard keyboard layout for Indian computers (from IS 13194: 1991)

~	! ₹	@	#	\$	%	^	&	*	()	-	+ =	Back	
1	2	3	4	5	6	7	8	9	0	Space				4th Row
Tab	Q औ	W ऐ	E आ	R ई	T ऊ	Y भ	U ड	I घ	O ध	P झ	{ [}]	\	3rd Row
Caps Lock	A ओ	S ए	D अ	F इ	G उ	H फ	J र	K ख	L थ	:	"	'	ENTER	Home Row
SHIFT	Z	X	C ञ	V	B	N ळ	M श	<	>		?	/	SHIFT	1st Row
SPACE														

Figure 15: Standard computer layout for Arabic computers

ز	!	@	#	\$	%	^	&	*	()	-	+ =	Backspace
Tab	ض	ص	ث	ق	ف	غ	ع	ه	خ	ح	ج	ب	
Caps Lock	ش	س	ي	ب	ل	أ	ت	ن	م	ك	ط	Enter	
Shift	ث	و	ر	ل	أ	و	ز	ط	Shift				
Ctrl	Alt								Alt				Ctrl

Figure 16: Early multiple shift computer keyboard for Japanese input (part).

スクリーン	テスト	S	R	Q
PA 1	PA 2	PA 3		
キャンセル	IDNT			E
印字	I	A		漢
FM	DNP			
確認				外

1	2	3
4	5	6
7	8	9
10	11	12

的敵適	倒凍唐	忒尼肉	舶薄麥	疲碑避	譜附武	便勉步	傍棒紡
笛摘滴	桃逃透	日入乳	爆縛箱	披罷美	部侮婢	保補捕	謀膨北
鐵迭哲	陶悼痘	尿任認	畑八発	備鼻尾	風封服	浦輔舗	牧睦撲
撒徹天	筒塔盜	妊忍寧	髮伐拔	微匹彦	副復福	母慕募	基朴僕
店点展	踏糖燈	年念燃	閔反半	筆泌姬	伏幅腹	暮簿方	堀本奔
転添田	騰鬪藤	粘納能	犯判坂	百表氷	覆払沸	包放法	翻凡盆
伝電殿	騰同動	農惱腦	板版飯	俵票評	仏物粉	報豊邦	麻摩魔
土徒都	堂童働	濃波派	帆伴班	標漂秒	奮紛憤	芳宝奉	磨每妹
吐途渡	銅導胴	馬婆拜	般販搬	苗描品	噴分文	朋胞峰	埋膜慕
塗努度	峠特得	配敗杯	範繁藩	貧浜賓	聞平兵	砲傲崩	又亦末
怒奴刀	徳匿督	肺背俳	煩頒番	敏不夫	陛丙並	訪飽縫	磨万満
冬当投	篤毒独	排廃輩	晩蛮盤	父付布	俳柄閉	俵防望	漫慢未
烏討党	突届豚	買売培	皮否肥	負富婦	弊米癖	暴亡乏	魅密稔
桐湯登	敦鈍曇	梅陪媒	非飛悲	扶赴浮	壁別辺	忙坊冒	脈妙民
等答統	奈内南	賠白博	費妃彼	善符腐	返変編	妨忘房	眠務無
灯頭豆	楠難軟	伯拍迫	批卑秘	怖敷膚	片偏遍	剖肪某	茅夢霧

One argument against this is that generally Asian languages use a greater range of letters than do European languages, so mapping is not always totally straightforward. However, this also offers the possibility of mapping QWERTY keys to a specific Asian script layout, so that Asian users need not learn the vagaries of the QWERTY layout if they are already familiar with the keyboard layout for their own language.

As stated above, there are national standards for keyboard layouts for some Asian scripts, but there are also many countries with no national standard. International standards may emerge over the next decade: the ISO/IEC subcommittee responsible for UNICODE has also recognised that it is necessary to be able to keyboard this multilingual repertoire, and one of its working groups already has keyboarding of the UNICODE repertoire as a new work item, although work has not yet developed much beyond that.

6. UNICODE and other character coding issues

26 The general issues of coded character sets for Asian scripts are well documented in greater detail elsewhere. and thus only general issues are covered below.

In order to reduce the cost of localizing products to provide foreign language versions of their software, manufacturers are already building in "globalization" or "internationalization" capabilities into the newest hardware and software. Computer architecture used to limit most character-sets to 8-bit sets, which in most cases limited a computer's functionality to English and at most one other script - and the coding for the other scripts varied greatly: in Figure 17 any 8-bit combination in columns A-F might represent . for example, either an accented Latin letter, a Greek letter, a Russian letter or an Arabic letter as well as an Indian letter depending on the application, the computer and the software in use: a situation prone to ambiguity when data is transferred globally. That limitation no longer applies. and current computers offer the capability of using 16-bit sets to code uniquely around 65,000 possible characters - enough to cope with all official languages and scripts in use worldwide. Figure 18 indicates unique codes for Devanagari which cannot be confused with other scripts.

Several large multinational companies, mostly based in the USA, decided that the open systems and systems integration approaches could be best served by agreeing a specific 16-bit character set for international use. This was developed by a group of computer manufacturers who formed the UNICODE Consortium for this purpose. and the coding in UNICODE is now identical to that used in the international standard ISO/IEC 10646:1991.

6.1 Limitations of UNICODE

Although UNICODE has the potential to allow computer use in every users language, just making appropriate alphabets and character sets available does not make the software immediately usable by those whose language is not English - the dominant language of computer applications. It will thus be some time before multilingual applications are developed for each language market. Even as it stands, UNICODE also has some limitations, in terms of the likely increased costs involved, and also in terms of the languages covered.

6.2 Increased costs

Although for Asian countries UNICODE provides the advantage of encoding most - but currently not all - scripts used in the national languages of Asian

Figure 17: Indian Script Code for Information Interchange (IS 13194: 1991)

HEX	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	
0	NUL	DLE	SP	0	@	P	.	p				ओ	ढ	र	.	EXT	
1	SOH	DC1	!	1	A	Q	a	q				औ	ण	ल	.		
2	STX	DC2	"	2	B	R	b	r				ऑ	त	ळ	.		
3	ETX	DC3	#	3	C	S	c	s				ः	क	थ	ळ	.	
4	EOT	DC4	\$	4	D	T	d	t				अ	ख	द	व	ी	
5	ENQ	NAK	%	5	E	U	e	u				आ	ग	घ	श	ी	
6	ACK	SYN	&	6	F	V	f	v				इ	घ	न	ष	ी	
7	BEL	ETB	'	7	G	W	g	w				ई	ड	न	स	ी	
8	BS	CAN	(8	H	X	h	x				उ	च	प	ह	.	
9	HT	EM)	9	I	Y	i	y				ऊ	छ	फ	INV	.	
A	LF	SUB	*	:	J	Z	j	z				ऋ	ज	ब	।	।	
B	VT	ESC	+	:	K	[k	{				ऐ	झ	भ	ि		
C	FF	FS	,	<	L	\	l					ए	त्र	म	ी		
D	CR	GS	-	=	M]	m	}				ऐ	ट	य	ु		
E	SO	RS	.	>	N	^	n	~				ँ	ठ	य	े		
F	SI	US	/	?	O	_	o	DEL				ओ	ड	र	.	ATR	

Figure 18: Uniquely coded Devanagari script in UNICODE\ISO-IEC 10646

00	FF	Hex	Letter
00	A zone (Alphabetic)	004C	l a n g u a g e . न ि प ी ।
10		0061	
20		006E	
30		0067	
40		0075	
50		0061	
60		0067	
70		0065	
80		002E	
90		0000	
A0	O zone (Open zone)	0000	
B0		0933	
C0		093F	
D0		092A	
E0		0940	
F0		0964	
	R zone (Restricted use zone)		

countries, there is an increased overhead. The main reason for this is the overhead in storage and transmission costs for many of its sources: in its official ISO form (ISO/IEC 10646) UNICODE requires 4 bytes for each character: even the permitted 2-byte option will still double the storage and transmission costs for all other languages compared to the norm of one byte per character outside of East Asia.

The implementations listed above involve relatively massive and expensive hardware to run UNICODE implementations. Ways will also be found to make UNICODE information more manageable given the prevalence of more modest computing resources in many enterprises.

For instance, Reuters in the UK is working on using UNICODE for indexing news items from its East Asian sources, in order to facilitate the development of a completely international and multilingual news service. from various parts of the world. However, Reuters stores the items themselves in their original form: for East Asian languages the 2-byte form will be incompatible with each other as sources in Chin.?.. Taiwan. Korea and Japan use different 2-byte codes for items in Chinese, Korean and Japanese. even though the same characters may be shared.

6.3 Range of languages

Currently there is only one commercially established implementation of UNICODE: Microsoft Windows NT. Although it provides the structure for all of UNICODE to be used, in the current version of this operating system only European languages (extended Latin, Greek and Cyrillic scripts) are supported, although Microsoft have promised a separate Arabic version.

An implementation for UNIX - the MASS system, developed in Singapore has been announced, which will cover Chinese and Tamil, as official languages of Singapore, along with English. Its developers also speak of the possibility developing versions with these languages for Windows 3.11 and Windows NT.

UNICODE itself also has some deficiencies as regards certain Asian languages. For instance the Indian scripts encoded have relied on outdated internal documents rather than national standards - i.e. ISCII (Figure 17). Using UNICODE coding means that only certain languages (e.g. English, Russian, Hebrew, Arabic) will get benefits such as direct sorting without developing additional subroutines, as they have used specific codepages which work that way - coding for other languages works out less fortunately.

27 Some scripts for national languages, and languages with several million speakers, are also currently not encoded. These languages include Burmese? Khmer, Maldivian, Mongolian. Sinhalese within South and Southeast Asia, and Tibetan and Yi within East Asia. A second edition of the international standard which encompasses UNICODE (ISO/IEC 10646) is being planned, and it is hoped that more contact will be made with the countries concerned than was done for the first edition. where outdated standards were incorporated.

28 Although the costs of two-byte transmission and storage will reduce over time applications, currently doubling the costs of these is a significant overhead. 8 bits satisfies most languages, except Chinese and Japanese (and Korean with hanja). 8 bit codes will easily provide for most Korean usage, in North and South Korea, despite the complexity of the compound shapes, which run into the thousands.

Chinese, Japanese and Korean computer users are generally happy with their existing two-byte systems and see little advantage in changing to yet another two-byte system invented in the USA.

6.4 Ways around the disadvantages of UNICODE

A 2nd edition of ISO/IEC 10646 will be developed before the end of the century. which is likely to remedy some of these deficiencies. The likelihood is that some computer systems will also develop in a UNICODE-compliant way. using 8-bit chunks of UNICODE rather than actually conforming to UNICODE in its full 16-bit mode. This approach may work in some large organizations and applications where they are well used to using a whole range of conversion routines. so one-byte to two-byte. or three-byte to two-byte conversion routines. should not add too many additional problems. The Windows environment will be increasingly important, both in a Windows 3.1 (S-bit character set. except for specific Chinese. Japanese. or Korean versions of software) and Windows NT (mostly always 16-bit character set) guises.

29 Some UNICODE-compliant developments in include UniType, developed by Gamma in Santa Monica, California. This is derived from word processing within a Windows environment. but will also work within other applications within a standard Windows 3.1 environment, but excludes Chinese. Japanese and Korean.

30 A similar text-based DOS equivalent. which is also capable of upwards migration to a Windows environment, is also being developed by SESAME Computer Projects of Harrogate, United Kingdom, which also provides a subset of the full UNICODE.

31 In India the GIST technology developed by CDAC in Pune is a graphical environment in both DOS. Windows and the Apple Mac environment, for most of the languages and scripts used in the Indian subcontinent. Although not extending much beyond South Asia and Southeast Asia, it has the potential for expansion beyond that area. Its advantage over UNICODE is in compatibility with 8-bit national standards, particularly for the Indian scripts - UNICODE is not compatible with the current Indian national standard.

7. Telecommunications and publishing

The widespread use of telecommunications in newspaper publishing predates the use of computers in the publishing industry by several decades. Telegraph. telex and telephone services were widely used among newspaper staff from the end of the nineteenth century, to increase the currency of news compared to the former reliance on physical transport systems. This increased the currency, but at the extra expense of involving additional staff to decode, rewrite. retype. or rewrite the original information, in comparison to the situation today where the reporter's original copy can, if required. be fed straight into the publication at the last minute.

7.1 Telecommunications in traditional publishing processes

The growth of the international news agencies led to the need to send data in several different languages. A number of bilingual solutions were developed, rather than a multilinear solution. to enable specific channels of communication between specific users, and different arrangements for each language might operate in different newspapers and in different news agencies. One of the more generally used approaches during much of this century was the Chinese Telegraph Code, which used four-digit numeric codes to represent specific Chinese characters out of the approximately 40,000 characters that might be used. This was also used in general commerce, as well as in newspapers. and required very highly trained encoders and decoders at each end. This was also used in some early computer applications as there were still people who used this as an input system. but from the mid-1980s more logical input and storage methods were used for Chinese.

Even today, most communication tends to be monolingual or bilingual, with standard uses set up - e.g. in India a newspaper might only want information in Hindi from a news agency which only dealt in information in Hindi, and the same might also be true of some news organizations in the Arabian Gulf. In both these cases two cases well established S-bit codes would be used. Even when news items are available from international agencies in the target language of the paper, quite frequently newspapers will get the information in the source language (typically English), and then translate the news item in-house into the target language of its readers. The large news agencies are having to change however, considering that both their news information and their customers are using a wide range of languages. For instance Reuters now employs someone fulltime to advise on UNICODE and its implications. as it regularly deals with information in Arabic and Japanese, as well as European language information. In terms of conventional publishing operations, multilingual use in telecommunications remains conservative. However, that may change during the next century, particularly when new information services and media based around computers start to become used much more heavily.

7.2. Telecommunications and new information services

The growth of both earthbound data networks, and of satellite communication has enabled countries to provide a wide range of new information services to their citizens. Some of the larger Asian countries have been in a position to provide their own satellites: others may see cooperation with other countries as the best approach to work in this field. Gulf Net is the information network which has been developed in the countries of the Arabian Gulf, which has been developed on a cooperative basis, and is widely used within the academic and business sector. An alternative approach is that of the Indian government, which is perhaps the most well-known Asian country to have launched its own communications satellites, which are used for a wide variety of broadcasting and news dissemination activities. The major advantage of satellites over earthbound communications infrastructures is simply that they require far less infrastructure. and the equipment that can be used is what citizens might buy for other reasons.

As an example, in India the number of televisions in use are twice as many as the number of telephones in use, and as a relatively standalone item television may be a very useful transmission medium for publishing. Again, as PCs are increasingly used in everyday business and home life in many parts of the world, CD-ROMS, and floppy discs are useful transmission media, for this relatively standalone equipment, as well as electronic mail, where the communications infrastructures exist. Computers are increasingly mobile, as are telephones, and battery power can enable mobile data transmission for both computers and telephones in certain circumstances.

7.3 A case in point: telecommunications in South Asia

As well as standard television programmes, the NIC and Doodardshan, the state television company, offer a teletext service called Intext, based on the well tried technique of transmission of data in parallel with normal broadcasts, but accessible to any television user with a simple decoder, and offering various information sources such as printed news summaries, and also nationwide railway and airline timetables. One specific development of the software has the potential to allow any of the services to display data in any one of the country's official languages, in any of the nine official scripts.

CDAC - the Center for the Development of Advanced Computing - is currently

embarked on a prototype programme for a further development of this approach. again enabling the use of any of the Indian languages, called Teletrak, with funding from the Department of Electronics of the Government of India. One advantage of using satellite transmission is that any part of the country can be reached, no matter how remote. As an example of this, satellite dishes mounted on trucks played a vital role in helping the planning teams, and enabling them to pass on vital information both to emergency workers and the local population during the earthquake in Maharashtra two years ago.

In terms of cooperation, there may be scope for some of the other countries in South Asia to benefit from similar approaches. or to cooperate directly with their larger neighbour, especially as their national languages work in a similar way to languages used in India. There are no technical barriers to this, although in some cases political considerations may make development slow in this area.

33 7.4 A case in point: telecommunications in Central Asia

The United Nations and some of its permanent programmes (such as the United Nations Development Programme - UNDP) and specialist agencies (such as the International Telecommunications Union - ITU) have assisted several Asian nations in providing satellite access. One of the most recent initiatives is benefiting Central Asian countries within the Commonwealth of Independent States (CIS). specifically Azerbaijan, Turkmenistan, Uzbekistan, Kazakhstan, Kyrgyzstan and Tajikstan, through UNDP's creation of country offices together with a satellite Wide Area network to strengthen communications among UNDP offices in the CIS and Eastern Europe. Although designed primarily for UNJDP use. it is intended to widen the audience to other agents of change, such as parliamentarians. government aid coordination agencies, chambers of commerce and the media. As well as offering access to databases of United Nations specialized agencies. it plans to extend access to databases in academic institutions. as well as to provide a secondary network using the worldwide Internet system. As well as intending to assist the region's integration into the world economy. it also has the potential to be a tool to increase public awareness.

Ventures like these have the potential both to assist in getting information to newspapers as part of' the general publishing process, as well as being a new medium of information to the end user, provided they have appropriate computing and communications facilities.

7.5. General barriers to multilingual telecommunications

However. communications is still one area where multilingual activity currently lags behind other developments. Some public access communication networks only allow 7-bit codes. Some data networks do allow the transmission of 8-bit, or even transparent data. but there are no universally accepted ways of automatically formatting data so that the data in the language concerned will be read by the receiver as it left the transmitter. UNICODE is only one of several ways in which data in a specific language can be transmitted, and costs twice as much to transmit as 8-bit encodings. By itself will not solve all these problems, until the majority of computer users worldwide use equipment that follows UNICODE. What is needed is additional work, and liaison, by organizations involved in standardization and telecommunications, at national, regional and international level. to ensure ways of doing this. At the international level, the ITU, and the Internationalisation working group within ISO are appropriate bodies to work on this: at regional level CPAL may be appropriate, together with the Asian Institute of Technology: at the national level, examples within the government of India are the NIC, RIND and CDAC.

8. Printing issues

34 The history of printing technology is discussed in section 3 above. As this relates to Asian scripts. Fiona Ross's book on Bengali printing provides a useful overview of trends up to and including the early uses of computers in printing. Many of these comments also apply to other Asian scripts. as parallel developments occurred these - not surprisingly as there are only three basic types of script (see Figure 1). However, computerization has made increasingly rapid strides, even within the last decade.

8.1 Technical processes

It is worth examining some of the technical processes involved, as regards languages. The processes involved with Desk Top Publishing (DTP) have had the greatest effect of publishing. since this computerizes nearly all pre-press processes. Originally developed with small-scale publishers in mind. the economies of scale they offered enabled existing companies to do much more, and new organizations to begin publishing.

35 The standardization of Postscript (a Page description language) and its adoption by many companies who provided and/or used printing technology, was a major breakthrough. Postscript used the page as the basic unit. and provided a means of controlling most processes - including text processes - as graphical entities. Some of the earlier systems were inflexible with editing once the page had been set up, as the text and columns needed to be placed around pictures. as well as taking into account surrounding columns, but more recent software is more flexible, and processes like hyphenation, kerning, and choice of font size on a page can be highly integrated to allow for last-minute changes. and the whole production process can be speeded up.

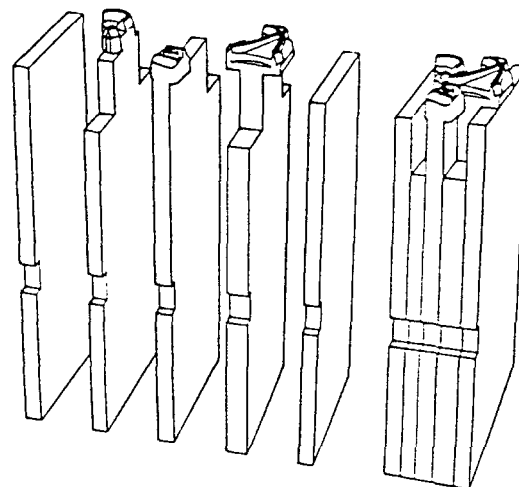
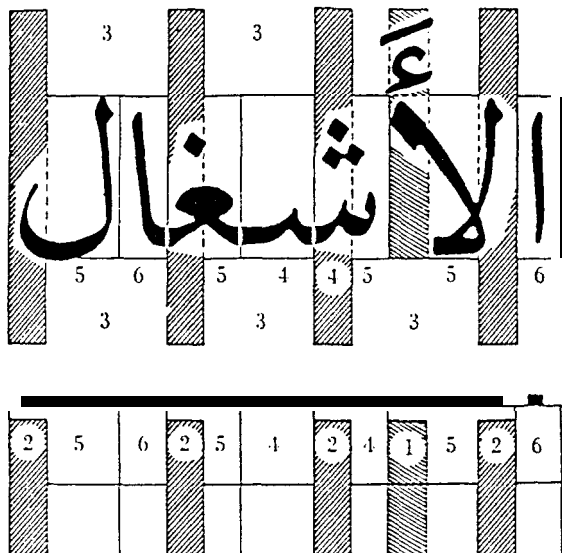
8.2 A case in point: kerning

Most scripts have letters of different width - the main exception being East Asian scripts, where characters usually occupy a square, and even here different spacing might be used to indicate vertical or horizontal direction of text. In Latin script the letters w and m are wider than i and l. Proportional spacing usually gives them the appropriate width in publications in order to aid reading. In most Asian scripts these differences are even more marked. The major developments in printing most Asian scripts came about once graphical handling of text replaced character-based systems in publishing - most Asian scripts could not be adequately represented on screen using - for instance - 8x16 pixel character cells, and even where the printer provided proportional spacing this meant that the screen looked considerably different to the printed version. The WYSIWYG approach available with most Macintosh, Windows, and even DOS publishing systems provide a much improved approach, and in many cases also allows any accent - or vowel sign - to be combined with any letter in most scripts.

As an example, some of the complexities of kerning (overlap of letters) are shown below. Figure 19 shows how this was dealt with in "hot metal" days - compositors would need to find the correct forms of letters, as well as to allow for possibly inserting vowel signs too. For running costs this represented a considerable operating and training overhead in terms of time, and in capital costs representing a considerable outlay to cover design and manufacture. However, nowadays this type of information can be held by the computer, and information on "kern spaces" used to calculate positioning of letters (Figure 20). Considering the range of scripts covered, and the use of scripts with similar properties elsewhere in Asia, India provides a particularly useful model in this area. Several Indian organizations, such as RIND and CDAC, and CALTIS have been active in printing developments here.

36 Figure 19: Using hot metal composition to achieve kerning in Arabic script

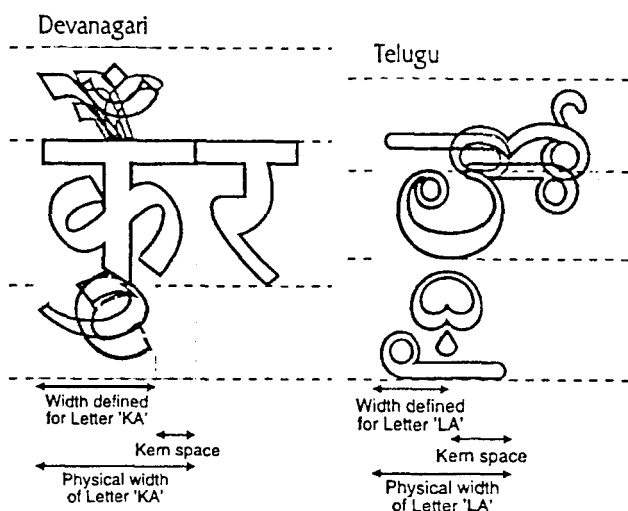
1. Unité de voyelle
2. Unités d'espacement des voyelles
3. Creux pour emplacement des voyelles
4. Caractère س découpé pour permettre l'emplacement d'unités de voyelle
5. Caractères avec crénage
6. Caractères sans crénage



L'illustration montre la haute complexité du système de composition mécanique voyellée présenté par la société anglaise Monotype. En haut une démonstration de l'assemblage des caractères mobiles, en bas les divers éléments intervenant dans la composition d'un mot, vue d'en haut et de profil. La lettre sin (4) se trouve découpée en deux pour permettre l'intercalage d'éventuels signes de voyellisation.

37 Figure 20: Using computer-held data to achieve kerning in Indian scripts

Kern Space Concept



Some complex compositions using Kern spaces:

Hindi	Malayalam	Telugu
दुथ	കമ്പ്യൂട്ടർ	సపై సైకర్
Oriya	Kannada	
ନିଶ୍ଚିତ୍ତ ପାରିପାଶ୍ଚିକ	ಸೂಚ್ಯಾಣು	

Alternate ways to represent the same word

विट्ठल विट्टल

9. Other issues: Retrieval, Sorting; Machine Translation, and Transliteration

Sorting and proper retrieval of information is important for most publications, whether it be to produce lists of personal names (e.g. for births, marriages, deaths. or for examination results) or of companies (e.g in stock market listings), or for annual indexes of articles, or for retrieval of information from newspaper archives for internal use. These issues are also important when new media are used, such as CD-ROM etc. as most of them rely heavily on structured indexing rather than merely free text to retrieve the appropriate data.

If items are not sorted as readers expect, they will complain if they cannot find information as they expect it. If reporters and researchers within the newspaper cannot find past articles and other information on a hot subject, there is a risk of current articles not being as precise as they might be,

The world increasingly relies on translation of everything from news items to large articles and books, from one language to another. The workload of translation agencies continues to grow. Computers play their part in these as much as any other organization, and in fact translation agencies provide a useful model for the take up of bilingual and multilingual technology that may be of some use to Asian language publishers. Transliteration may also affect how effectively information is retrieved in some circumstances.

9.1 Retrieval

This last situation is a real one. The research librarian of an Arabic newspaper recently described to me how their computerised, Macintosh-based, indexing system failed to locate certain past articles because the information had been input correctly. but some included hamzated letter forms (in effect accented forms) while on other occasions the plain letter forms of the same term had been put in. Without dealing with this, some archive material could easily become lost.

European language computer systems are nowadays expected to enable retrieval of the same terms whether or not the accents are used: the same criteria ought to apply equally to Asian language terms in computer systems. However, all too often this is not apparent until the system is installed and paid for, and extra costs are incurred to put it right, or else it may not be possible to correct this fault.

Misspelt terms will always get lost, but even here it is increasingly possible to pass lists of terms through a spellchecker: it is important that all critical software used in the publishing operation is sufficiently language-aware.

9.2 Sorting

It is important to ensure that any software acquired allows sorting to take place in the proper order. and in some cases special software may be required. Even in UNICODE. which has been designed to allow the use of all Asian scripts in current use, does not provide direct sorting without intervening software, except for a few languages, specifically English, Greek, Russian, Hebrew and Arabic.

All other languages, including those which use extended versions of these scripts, will need additional software. Most bilingual 8-bit operating systems do provide for this without the need always to use special sorting software, but such bilingual systems do not exist for all Asian scripts.

In some cases, computer manufacturers must still wait for sorting orders to be established before they can provide appropriate software: an example is Mongolian, where different sorting orders are in use in both the Inner Mongolian Autonomous Region of China and in the Mongolian Republic.

9.3 Machine translation

38 One feature of translation is that mother tongue speakers produce far more readable copy in the target language than do mother tongue speakers of the source language. A similar consideration also applies to machine translation. This is still, in its early days as regards commercial systems, although much government funded research in the West has gone on for the last two decades. No commercial machine translation systems exist for Asian languages, except between English and Japanese; however machine translation research has been undertaken for several years in Tunisia on developing systems for translation to and from English and Arabic.

It is likely that even with commercial systems available from companies like IBM (with its Translation Manager 2), even for European languages machine translation will mean machine assisted translation for many years to come. Even then, such is the variability of use of languages that it seems unlikely that machine translation will ever reach mother tongue translation status, except for highly specialist use, with limited vocabularies - weather forecasting systems were one of the few successes in the early days. The old apocryphal story from these times is likely to be as good as ever: how do you stop a computer translating the English phrase "out of sight: out of mind" as "invisible idiot" in the target language?

9.4 Machine transliteration

39 One area which may be of more immediate use to publishers in some respects than machine translation is machine transliteration. This would enable consistent use of terms, particularly names, to be used both in publishing, and in archiving, and retrieval - particularly where personal names and place names are concerned. The TRANSPARENT project in India is an example of this, where names in directories can be automatically displayed in any Indian script, and also in Latin letters.

This approach is particularly applicable in countries where the same, or closely related languages, are used with different scripts. Indexing terms in a common script and allowing output of terms in other scripts would also save effort. The former Yugoslavia provides the most well known example in Europe of where this approach was used - the same words could be represented in either Cyrillic script or Latin script on an almost exact one-to-one basis. Examples of where this could apply in Asia include India, where place names in particular are pronounced the same in both Hindi and Urdu, even though represented in different scripts; similar considerations apply, in proper names at least, in Hindi and other official languages/scripts of India.

Computers could also help in other countries in Asia, particularly where active changes of script is taking place, or likely to take place. Examples are in Azerbaijan, and possibly later the other Central Asian republics within the CIS, where the script is changing from Cyrillic to Latin script; and also in Mongolia where the script is being changed from Cyrillic to Mongolian script.

10. Looking beyond the basic issues

In publishing, as in all walks of life, it is all too easy to become immersed in the technical issues and everyday events that make up most of our working life. The nature of publishing however, is to look outwards to the wider world, both for the producers and suppliers of information and for the readers.

The nature of today's world is shaped by global events as much as national events. One of these global factors is the increased amount of business done between nations, and an increasing amount of travel and migration among countries. Thus each country has to take account of large numbers of people speaking different languages: some countries, especially the larger countries such as China and India have had to deal with this situation for many centuries. Publishing in these countries has grown up with these realities, and it may be useful for countries increasingly facing this situation to learn from these two, different, Asian models, as well as other parts of the world where this situation has come into being more recently.

The key to all this is cooperation at all levels, and planning and acting together, and that is the focus of part 3 of this background paper.

Part 3 - Synthesis of basic issues and recommendations

11. Synthesis of basic issues

11.1 Overview

Although the computer revolution, with its massive impact on publishing, began in the West, changing economic conditions mean that Asia may begin to have certain advantages particularly because of the vast wealth of talent in computer applications in newly industrialized Asian countries, and the competitively priced skills compared to costs of similarly trained staff in the West. This could enable the traditional East-West dependency to be reversed. In terms of Asian publishing this would enable:

40

- (a) the defining of Asian rather than Western priorities, and
- (b) in being able to supply Asian solutions to meet Western needs, and exporting Asian technology to the West, and or enabling inward investment.

11.2 Prime issues concerning languages

In terms of script developments, governments need to ensure that any plans for language reform are well known, and carried out to a realistic timescale. Script usage in most languages seem fairly well settled, although specific actions are already taking place, or likely to take place soon, in parts of Central Asia, Mongolia, and Hong Kong after 1997. Any reforms also need to feed into all parts of the education system, so that there is a pool of potential employees who can work in the publishing industry, and also of readers for their publications.

Computer suppliers need to ensure that their equipment meets the language needs of the publishing industry with a full and appropriate range of characters, and with appropriate printing, telecommunications, sorting and other facilities for these languages. Computer companies, and standards organizations need to ensure that the standards, equipment and software that they produce is based on actual Asian needs, and not on Western perceptions of Asian needs. For some languages (particularly in Southeast Asia) issues like word division - i.e. where there are no spaces - may cause additional problems which need solutions, e.g. in hyphenation, and in indexing.

11.3 Prime issues concerning computers

Considering the large investment in equipment, software, support and skilled manpower involved in publishing technology, it is worthwhile for publishers to employ, and/or to buy in, professional assistance to cope with this rapidly changing technology. No-one wants to find that their equipment has suddenly become obsolescent, or that companies providing systems or support have gone bankrupt. Knowing what choices are available, and what are the effects of specific actions is important to the viability of any enterprise.

Although microcomputers are the front end to most publishing systems, a publishing system is more than just the sum of its constituent parts. Providing suitable systems integration, and dealing with any "legacy" systems which may be out of date, but still have some usefulness left, and relating the publishing system to other computerised operations in the organization (such as purchasing, payroll etc.) may also make economic sense.

There are also a wide variety of technical issues, such as interfacing different computers, operating systems, and devices, and ensuring that data is consistent throughout, and beyond, the organization. There is also the issue of dealing with new computer media, as well as of interaction with broadcasting organizations, as news sources are shared, particularly as there

is increasing common ownership by the same company of both broadcasting and publishing enterprises.

Some traditional printed publications are also using new alternative media, such as online information sources, and CD-ROMs, to provide something nearer the immediacy of broadcast media for rapidly updated information, especially for structured information in tabular/database forms. Again, there are rival computer platforms and operating systems: which will dominate is unsure, although it is likely that future operating systems will offer some degree of compatibility with existing operating systems.

Several concerns, both current and future, revolve around input issues.

41 Future input issues include voice input and OCR (Optical Character Recognition). Both are in their infancy, even for most European languages, and especially as regards commercial systems for most Asian languages. Some information sources exist already but need inputting again as they are in a different language. Machine translation (MT) may improve over the years but there will still be a major reliance on human translation, even if MT becomes more widely used.

Asian language keyboard layouts still need sorting out by manufacturers and standards bodies alike. West Asian and East Asian keyboard layouts are the most standardised, but several languages between these areas use a variety of input systems, none of which is prevalent, or even optimised for use by typists in those languages. Physical keyboard design to reduce RSI may also produce some new approaches in the next few years.

11.4 Standardization requirements and realities;

Keyboard layout will be increasingly looked at by standards bodies, but given the variety of Asian languages, character coding will be one of the big issues over the next decade. Most advances are likely to be built up around the 16-bit UNICODE, but not all Asian language issues have been addressed - coding for some scripts is inconsistent with that of national standards, and not backwards-compatible with existing 8-bit standards, and others have not been coded. Standards bodies also need to recognise that the use of UNICODE means increased storage and transmission costs for many languages.

The Internet, and other electronic transmission media will form an increasing path for some published data. However, these worldwide networks will need to improve methods of handling multilingual data, and sending it by the transmitter, and appropriate decoding by the receiver, whatever character set is involved by each.

11.5 Conclusions on the impact of technology in Asian publishing

Far Eastern computer systems are already dominating Western markets, and an increasing number of Western companies are already locating much of their computer R&D work to India, so Asian countries are now in a much stronger position to make use of new technology in publishing. Asian countries can capitalise further on this through cooperation at all levels between organizations in each country, whether individual publishers, news agencies, printers, typographers, distributors, computer organizations, etc., either individually or through appropriate trade associations and professional associations; through national, regional and international bodies.

Once the goal has been established, of enabling Asian language publications to reach many more of their people, in ways that make the most of available technology, and relevant focal points established, this goal can be pursued in every national, regional and international forum. This will provide a shining example of using technology to serve mankind.

12. Recommendations for action

Section 12 provides a series of suggestions in a variety of areas, which could be used as a basis for formal recommendations during the conference. It looks at national cooperation, which can be done by governments, the prospects for regional cooperation within Asia, and for international cooperation, particularly pointing out a few of the ways in which Unesco has been involved in the past, and could become involved in the future. It also provides some suggestions for publishers and individuals.

12.1 Government action:

42 Governments should ensure that they have active standards bodies coordinating activities regarding use of their national language on computers. They should ensure that they are represented on the International Organization for Standardization (ISO), Concrete examples are India, which plans to upgrade its participation in the International Organization for Standardization (ISO) from Observer (O-member) status to participant (P-member) status, and has sent a delegate to working group sessions of ISO/IEC-JTC1/SC2/WG2 and SC2/WG3; and China, which has also ensured that these ISO working groups are aware of all its minority languages and the scripts that they use, so that these can be included in ISO 10646. The recently announced plan by the European Computer Manufacturers Association (ECMA) to "fast-track" a new edition of ISO 10646, which would also encode those Asian national scripts not already encoded, may in fact be an additional opportunity for Asian countries, rather than a European imposition. Cooperation with ECMA may be less costly than with ISO, and resulting standards are widely accepted as ISO standards, and currently ECMA standards are available free of charge. Asian governments should ensure that their standards bodies are in contact with ECMA regarding developments with their scripts.

It is also often up to governments to ensure that adequate telecommunication links exist, whether through terrestrial means or via satellite. Many Asian governments also run the major news agencies in their countries, and government regulation may determine the environment for the development of other news agencies. Governments can ensure that appropriate technical know-how is available to news agencies and newspapers, so that the free flow of information is not subject to any limitations of language technology.

12.2 Regional cooperation - assessment of existing organizations

Through Asia as a whole, Unesco has played an important role in helping promote regional cooperation. Examples are the two CPAL conferences (Computer processing of Asian languages) co-sponsored by Unesco. The Asian Institute of Technology (AIT) in Bangkok, has acted as a secretariat for CPAL, and as a clearing house for relevant information in this field, having been involved for many years in automation projects involving Asian languages. I recommend that Unesco also contact other relevant centres in Asia, such as CDAC in Pune, which are already involved in assisting in publishing ventures, and ALDOC which plays a major regional role among Arab countries, and Unesco is in a good position to identify other possible focal points too.

As the use of scripts sometimes transcends national boundaries, it would also be useful to encourage regional cooperation at a more local level. For example, where there are well established regional intergovernmental organizations in specific parts of Asia, such as the Association of South East Asian Nations (ASEAN) and the Gulf Cooperation Council (GCC), these may be appropriate bodies to foster cooperation and specific projects. In other cases, regional trade associations may be more appropriate.

12.3 International cooperation: roles for Unesco and other organizations

Unesco and its national commissions have already had an influence in the area of language technology through work it supports. Examples are: (a) the Workshop on computerization of Burmese, Khmer and Lao Scripts, held from 15-17 July 1993; (b) the wide-ranging 1993 Unesco survey on Khmer-font coding by Maurice Bauhahn; (c) assistance provided by Unesco and others to typographers in Asia, e.g. through the provision of experts to help develop Thai and Lao typography in the late 1970s; (d) its development of multilingual software tools such as CDS/ISIS, in use in many organizations throughout Asia, mostly in an MS-DOS environment, which could well be useful in archive work in publishing enterprises; and (e) Unesco's liaison status with the International Organization for Standardization (ISO).

There are further specific things which Unesco could do in this area.

(1) It may be useful for Unesco's liaison status could be used to provide expert representation of the needs of Asian nations to the appropriate standards working groups concerned. Unesco has not sent representatives to the appropriate committees in recent years, and some of the Asian countries concerned been in a position to send them. These actions might help ensure that the full range of Asian scripts are available in future equipment.

(2) Further assistance in specific projects relating to language technology may also be appropriate, along the lines of those listed above, particularly where the commercial marketplace had not already provided adequate solutions.

(3) In particular, considering the usefulness of CDS/ISIS as a tool (see (d) above), and its widespread use. it is essential for Asian users that Unesco's recent development of a Windows version of its search-engine for CDS/ISIS is "language-aware" for the particular characteristics of various Asian-language versions of Windows.

Several organizations, both regional and international, cover the fields of computing and publishing and related issues discussed in this document. Identifying issues and priorities for action in this area by IGOs and NGOs and coordinating and funding any useful activities may be important where local conditions may mean that the private sector or national governments are not in a position to take a lead. Examples of such organizations include the International Federation of Master Printers, the International Institute for Journalism in Berlin, the Graphic Arts Foundation, and the Graphic Media Development Centre in the Hague (formerly part of the State Printing Office), IFRA and INES. In some cases national organizations (such as CDAC, RIND and CALTIS) may be in a position to offer regional assistance beyond their national boundaries.

It would be useful to compile a directory of such organizations! and to involve them in joint action in this area. Coordination with other organizations within the United Nations family of organizations, such as UNDP (United Nations Development Programme), the World Bank, and the ITU (International Telecommunications Organization) would also be useful, particularly where funding may be available. Identifying good practice and model setups in Asia, and enabling visits to them by appropriate technical staff could also be useful (see also section 12.4) below.

12.4 Action by the private sector

When acquiring new equipment, publishers should push for appropriate solutions for ASian language publications, e.g. to ensure the availability of widely accepted keyboard layouts for specific Asian languages, and are likely to take care of future needs, such as increased use of telecommunications and new media.

Suppliers should also ensure that equipment supplied is able to take care of local needs in developing countries. for instance to allow for interruptions in power supply. Building in solutions like the battery-powered Robust PC, developed in India. and the use of laptops for certain tasks, as well as the use of reliable in-house generators, need to be considered by companies providing integrated systems.

It would be useful for many Asian companies to arrange study visits by technical staff. both to other companies and organizations both in Asia and the West, Many technical developments still originate in the West. although the balance is beginning to change due to the cost advantages that Asian countries can offer, as pointed out in section 11 above. Many companies and experts from the newspaper publishing industry worldwide are based in and around London and other Western capitals. It may be useful to plan meetings of NGOs in these capitals who are involved with developments relevant to newspaper production in Asian languages (see also 12.3 above).

One benefit of visits to or meetings in Europe is that several major companies, organizations and experts which could provide expertise to Asian countries are located near together. Advantage could also be taken of the fact that such organizations can afford to participate in events like the current conference, and indeed may see commercial advantages in doing so, and even in sponsoring some activities. This would allow more of Unesco's funds to be channeled towards sponsoring travel for Asian participants.

For example, meetings in London might involve participation from Linotype and Monotype (as developers of multilingual typesetting equipment with perhaps the largest installed base of equipment worldwide), Diwan (to cover some of the West Asian language aspects. and aspects of systems integration): East Word/South China Printing Company (to cover some of the East Asian language aspects); Reuters (to cover news gathering elements and data transmission); Seybold Limited (who provide a series of international technical newsletters on various aspects of newspaper publishing, and who have regularly covered non-roman script automation in their Seybold Report on Publishing Systems); Maltron, (covering issues such as keyboard layouts, ergonomics and health and safety) as well as several national and international newspapers (including the Arabic Al-Hayat and the Urdu Daily Jang) and a wide range of consultants and software suppliers in the area of Asian language systems. Similarly useful participants could probably be found in other European capitals.

12.5 Action by individuals

Whatever the action of governments, regional and international organizations, or publishers, and whatever mechanised or computerised systems are installed. the success of any newspaper depends on the people in it.

New technology can be seen as a threat when existing practices and procedures. with which people have sometimes undertaken for much of their life, when changes are introduced suddenly. Computerised systems need to be understood as a servant to help undertake a task, and which can be made to fit into the ways people work, not new masters to which people are forced to conform. Changes in procedure often come about only gradually, when people suddenly spot ways in which they can do things better.

Given the increasing power and availability of computers, and the way computer designers are beginning to address Asian language needs as much as other users, and their increased use within the newspaper publishing industry, new technology can provide new opportunities, and how people use its potential can be a major key to success.

REFERENCES

- 1 For an indexed list of scripts used for Asian languages, see CLEWS, John A classification of scripts SESAME bulletin 5 (2) Summer 1992: 63-74.
For a historical overview of scripts. see also DIRINGER D The alphabet. 1968, 2 volumes.
- 2 Source of diagram: CLEWS. John Standards for the living languages of the world In: Akshara '94: the International Conference on Applications of Information Technology in South Asian Languages, edited by S.S. Agrawal and Manish Jain. New Delhi: BPB. 1994. [In press].
- 3 Source: the "Big 5" Chinese character set mainly used in Taiwan.
- 4 INTERNATIONAL PHONETIC ASSOCIATION. The International Phonetic Alphabet, revised to 1979. London: International Phonetic Association, University College London. Department of Phonetics, 1979.
- 5 JONES, Daniel. The problem of a national script for India. Lucknow: Pioneer Press: Hertford: Stephen Austin and Sons, 1942.
- 6 WESTERMANN, D and TUCKER, A N. A practical orthography of African languages. London: International African Institute. 1930.
- 7 ROMAN LIPI PARASHAD Press release, 10 July 1984. Bombay: Roman Lipi Parashad, 1984.
- 8 CENTRAL HINDI DIRECTORATE. MINISTRY OF EDUCATION, GOVERNMENT OF INDIA. Standard Devanagari script. New Delhi: Central Hindi Directorate, 1966. (Publication no 3/67).
- 9 INDIA. Central Hindi Directorate. Devanagari: development, amplification and standardisation. Central Hindi Directorate, Ministry of Education and Social Welfare, Government of India, 1960.
- 10 HAMM, Roberto Pour une typographie arabe: contribution à la démocratisation de la culture arabe. Paris: Sindbad. 1975. (La bibliothèque arabe: collection Hommes et société)
- 11 Great Walls tumbling down? SESAME bulletin 4 (4) Winter 1991: 120-121. Summary of: LIU Yanwen, GARDI and ZHANG Zhu The Mongolian, Chinese and English operating system. Chinese journal of computers 13 (2) 1990: 124-132.
- 12 CPAL-1: Computerizing Asian languages, Bangkok, September 1989 SESAME bulletin 2 (3-4) Summer-Autumn 1989): 98. Also in Unesco reports.
CPAL-2: SINHA, R.M.K., ed. CPAL-2: Computer processing of Asian languages: proceedings of the second regional workshop. New Delhi: Tata McGraw-Hill, 1992. ISBN 0 07 462422 9.
- 13 Following the Design information report on text composition in Devanagari, by S.P Mudur, L.S. Wakankar and P.K. Ghosh, produced by the National Centre for Software Development and Computing Techniques at TIFR in Bombay, The Research Institute for Newspaper Development in Madras produced Design information report[s] on text composition in other Indian scripts, such as Telugu and Malayalam.
- 14 Several guides to Asian language software have been produced. Some of those tend to concentrate on the Apple Macintosh, for instance: SCHNEIDER, Seth, ed. Multilingual computing. Clark Ford, ID: Worldwide Publishing, [1993].

MACDOUGALL, Susan Non-roman scripts: a software directory.
Canberra: S. McDougall. 1988. ISBN 0 9599120 3 7.
GANLY, Tony; GILDING, Jack and GREEN. Peter Multilingual word processing.
Melbourne: Community Technology, [1989?].
VENKATACHARYA, Patañjali S. Sandhi: the guide to multilingual computing.
Toronto: ICANAS, 1990. ISBN 0 9694770 0 7.

Is of books which concentrate on the PC, Ian Tresmans's directory listed below is the most comprehensive and the most frequently updated by new editions. Updates have also appeared in SESAME bulletin since Spring 1992. TRESMAN, Ian The multilingual PC directory. Elstree: Knowledge Computing. 1991, and subsequent editions.

- 16 Several Apple publications document Asian script processing. including:
APPLE COMPUTER. Guide to Macintosh software localisation.
Reading, Mass: Addison Wesley, 1991. ISBN 0 201 60856 1.
APPLE COMPUTER. Inside Macintosh, volume V. Reading. Mass: Addison Wesley, 1986. ISBN 0 201 17719 6. Chapter 16: International utilities package;
Chapter 17: Script manager.
APPLE COMPUTER. Inside Macintosh, volume VI. Reading. Mass: Addison
Wesley, 1991. ISBN 0 201 57755 0. Chapter 14: Worldwide software overview;
Chapter 15: TextEdit
APPLE COMPUTER. Inside Macintosh: Quickdraw GX: typography. Reading. Mass:
Addison Wesley, 1994. ISBN 0 201 40679 9.
- 17 BRUNNER, Laurel Diwan and broadcast printing: Arabic papers for 10 cities
Seybold report on publishing systems 20(1), September 3, 1990: 15-23
See also O'KELLY, Sebastian The Arabian news: case study DTP: corporate
publishing and presentations April 1989: 27-28.
- 18 Paperless Pre-press Gist Bulletin (5), March 1994: 34-35
- 19 See for instance LEDBETTER, James Reporters' injuries blamed on video
terminals New York Observer, December 25 1989 - January 1, 1990: 9, and
several other recent cases involving newspaper staff.
- 20 Many articles papers have been written by Stephen W. Hobday on the Maltron
keyboard. The diagrams in Figure 10 are taken from:
HOBDAY, Stephen W. A keyboard to increase productivity and reduce postural
stress. Trends in ergonomics: proceedings of the Annual International
Ergonomics and Safety Conference: edited by Fereydoun Aghazadeh.
Amsterdam: Elsevier, 1988: 321-330; and from
HOBDAY, Stephen W. Computer related upper limb disorder: a keyboard to
eliminate the stress and the pain: an interim success report. Proceedings:
19th Annual Congress of IMART, 4 May 1994. London: The Royal Society of
Medicine, 1994.
- 21 Taken from: DVORAK, August There is a better typewriter keyboard National
business education quarterly 7(2), December 1943; 51-58 and 66.
- 22 For a variety of typewriter keyboard layouts, both Asian and European see
BEECHING, Wilfred A. Century of the typewriter London: Heinemann. 1974.
ISBN: 0434 90120 2.
- 23 CENTRE FOR DEVELOPMENT OF ADVANCED COMPUTING. INSCRIPT keyboard:
Indian scripts rediscovered. Pune: Centre for Development of Advanced
Computing, 1993: 1-36 [diagram: page 4].

- 24 MICROSOFT CORPORATION Microsoft MS-DOS operating system version 3.30: User's guide. Arabic supplement. release 1.2. Redmond, WA: Microsoft Corporation, 1989; C-3.
- 25 CHIN, Wey-tze Chinese language automation in libraries
SESAME bulletin. 6 (1), Spring 1993: 12-18.
- 27 Unicode technical reports 2 and 3 actually provide suggestions for coding these scripts. as did the Unicode Version 1.0 preprint of April 1, 1990. However, the Unicode Consortium has funded additional research to enable proposals for these scripts. but it is the ISO/IEC which is responsible for allocating the coding. Neither the Unicode Consortium nor ISO/IEC have made much contact with national standards bodies in Asia unless the standards bodies have approached them first. The Unicode Consortium has also been very slow to provide ISO/IEC with details of these proposals.
- 28 KIM, Kyongsok A common approach to designing the Hangul code and keyboard. Computer standards and interfaces, 14 (1992): 297-325.
For information on S-bit coding in North Korea. see CLEW'S, John
A Korean version of ISO 8859? SESAME bulletin 4 (2), Summer 1991: 52-54.
- 29 GAMMA PRODUCTIONS Gamma Unicode server, version 2.0: preliminary description. September 27, 1993. Santa Monica, Gamma Productions, 1993.
- 30 SESAME COMPUTER PROJECTS. SESAME Computer projects announces multilingual DOS fonts and drivers [press release]. Harrogate: SESAME Computer Projects. 1994. Since then, coordinated developments with other software developers have led to the prospect of similar script offerings across Windows and Macintosh environments as well as for DOS.
- 31 Details of CDAC's software and hardware developments for Indian scripts can be found in GIST bulletin (Pune: CDAC, 1992-).
- 26 See for instance:
CLEWS, John Language automation worldwide: the development of character set standards. Harrogate: SESAME Computer Projects, 1988.
CLEWS, John Standards for the living languages of the world In: Akshara '94: the International Conference on Applications of Information Technology in south Asian Languages. edited by S.S. Agrawal and Manish Jain.
New Delhi: BPB. 1994. [In press].
See also articles in the Standards News section of SESAME bulletin.
(Harrogate: SESAME Computer Projects, 1986 -) and its successor Global Word.
- 32 National Informatics Centre, TDPP, 2nd Floor, A-Block, CGO Complex, Lodhi Road, New Delhi - 110003, New Delhi, India. (tel: 4360855: fax: 4360597). A variety of newsletters and brochures are available from the NIC in this area.
- 33 Update (UNDP), 6(24), December 1993: 1-2
- 34 ROSS, Fiona The evolution of the printed Bengali character from 1778 to 1978. London: School of Oriental and African Studies [SOAS]. 1988. [Ph. D. thesis: due for publication by SOAS, 1995]. This provides a historical overview which mirrors similar developments for other scripts in Asia.
- 35 ADOBE SYSTEMS INC. Postscript language reference manual. Reading, Mass: Addison-Wesley, 1985. ISBN 0201101742.

- 36 Source: HAMM. Roberto Pour une typographic arabe: contribution à la démocratisation de la culture arabe. Paris: Sindbad, 1975.
(La bibliothèque arabe: collection Hommes et société)
- 37 CENTRE FOR DEVELOPMENT OF ADVANCED COMPUTING. The publishing revolution: CDAC - GIST - ISM [ISFOC Script Manager]. Pune: Centre for Development of Advanced Computing. 1994: 5.
- 38 See for instance: Machine Translation Summit [I. Japan. September 1987] ed. by M. Nagao [et al.] Tokyo: Ohmsha. 1989: also Machine Translation Summit II: Munich. 1989. and subsequent meetings in this series. Work on Arabic machine translation has been done in Tunis - see Arabic computing developments in Tunis SESAME bulletin 4 (4), Winter 1991; 121-122. Initial work has also been undertaken in India through the Department of Electronics.
- 39 TRANSSNAME: the Names conversion software GIST bulletin 2(1). February 1993: 10-11
- 40 A number of these issues are explored in some of the following publications: SUBRAMANIAN, C.R. India and the computer: a study of planned development . New Delhi: Oxford University Press, 1992. ISBN 0 19 562 735 8: SINGH, Arvind and ROSS. Everett M. India's information revolution. New Delhi: Sage. 1991. 0 839 9616 0: Information technology manpower issues for developing countries / editor S.C. Bhatnagar: New Delhi: Tata McGraw-Hill, 1992: and Social implications of computers in a developing world / editors: S.C. Bhatnagar, S.E. Mayuri Odedra. - New Delhi: Tata McGraw-Hill, 1992
- 41 A number of OCR prototype projects have been developed for several Asian scripts. Examples include: SINHA, R.M.K. and MAHABALA. H.N. Machine recognition of Devanagari script IEEE Transactions on Systems, Man and Cybernetics, 9 (S). August 1979; 435-441; and SETHI, I.K and CHATTERJEE. B. Machine recognition of constrained hand printed Devanagari. Pattern recognition 9, 1977: 69-75.
- In the mid 1980s, the Sanskrit Text Archive project was faced with a major amount of keyboarding in Sanskrit, and OCR was examined as an option at that time, but the technology was still not sufficiently advanced. See: WUJASTYK, Dominik The Sanskrit Text Archive Conference SESAME bulletin 2 (3-4) Summer-Autumn 1989: 92-95.
- There have been three Arabic OCR systems developed in various parts of the world, but none are yet a major commercial success or without significant recognition problems compared to Latin script equivalents.
- 42 Examples of these scripts can be seen in any issue of Minzu tuanjie. (Beijing: Minzu Chubanshe). Also reprinted in CLEWS, John. Language automation worldwide. Harrogate: SESAME Computer Projects. 1988. For a limited period the Chinese title Minzu tuanjie was also translated into Zhuang (reman script), Tai Nua and Yi.
- 43 WUWONGSE, Vilas Workshop on computerization of Burmese* Khmer and Lao Scripts, 15-17 July 1993: [report]. Vilas Wuwongse Computer Science Program, School of Advanced Technologies, Asian Institute of Technology, Bangkok, Thailand
- 44 BAUHAHN, Maurice - Unesco survey on Khmer-font coding [Phnom Penh]: Unesco Cambodia, 1993. 38 pp.