The Variety of Scripts and Reading

Insup Taylor
The Variety of Scripts and Reading

Insup Taylor
Life Sciences, Scarborough Campus
University of Toronto, Canada

A variety of writing systems are used today around the world. They differ in appearance, number of symbols, and above all, in the linguistic units coded. In this address, I will describe four writing systems and 10 of their extant varieties or scripts, and then discuss how they are learned. Table 1 presents an overview of the writing systems and scripts to be described.

Table 1. Variety of Writing Systems and Scripts

<table>
<thead>
<tr>
<th>System</th>
<th>Script</th>
<th>Lng. Unit</th>
<th>No. of Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logography</td>
<td>colored shape</td>
<td>word</td>
<td>130?</td>
</tr>
<tr>
<td></td>
<td>Blissymbol</td>
<td>word</td>
<td>500?</td>
</tr>
<tr>
<td></td>
<td>Chinese character</td>
<td>morpheme</td>
<td>50,000?</td>
</tr>
<tr>
<td>Syllabary</td>
<td>Cree-Eskimo</td>
<td>syllable</td>
<td>44-48 (+ ?)</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>syllable</td>
<td>210 (+ ?)</td>
</tr>
<tr>
<td></td>
<td>Japanese Kana</td>
<td>syllable</td>
<td>46 (+ 60)</td>
</tr>
<tr>
<td>Alphabet</td>
<td>English</td>
<td>phoneme</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Finnish</td>
<td>phoneme</td>
<td>37 (+ 6)</td>
</tr>
<tr>
<td></td>
<td>Hebrew</td>
<td>consonant</td>
<td>22 (+ 9)</td>
</tr>
<tr>
<td>Alphabetic</td>
<td>Korean Hangul</td>
<td>phoneme</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>syllabary</td>
<td>syllable</td>
<td>2,000?</td>
</tr>
</tbody>
</table>

In Table 1, under "No. of Symbols," a question mark indicates that the number is unspecifiable, for reasons to be given under each script. A logography codes meaning unit, while a syllabary, alphabet, and alphabetic syllabary codes phonetic units. Historically, writing systems developed in this order.

I will consider three themes:

1. The characteristics of a language dictate, to some degree, the type of script preferred by that language.

2. The way a script represents meaning and/or sound has a great influence on how word recognition in that script is learned.

3. However, a script is not the sole influence on reading achievement and literacy attainment; socio-economic factors also play a role.
Logography

Each logograph represents a small meaningful unit, a word or morpheme. It may not be associated with any sound, or it may be associated with more than one sound. Consider one kind of logograph, the Arabic numeral 4. Its meaning remains constant, even to a deaf-mute who cannot read it aloud, or to speakers of diverse languages who sound it out variously as *four* (English), *quatre* (French), *yotsu* or *shi* (Japanese) and *net* or *sa* (Korean). In this sense, a logograph represents meaning primarily and directly and sound only through the meaning.

For rudimentary "reading and writing," a handful of logographs are adequate. Two such examples are colored shapes for chimpanzees and Blissymbols for cerebral-palsied children. For proper reading and writing, a full logography, a system of logographs, is necessary. The Chinese system is a prime example of a logography.

Colored shapes for chimps

The simplest logographs may be the colored plastic shapes devised by the psychologists Premack and Premack (1972) to teach chimpanzees to communicate with human trainers. The plastic shapes are backed with metal so that they adhere to a magnetic board. Each arbitrary colored shape represents one word, and a few such shape-words are mounted on a magnetic board to make a "sentence," as shown in Fig. 1.

![Figure 1 — Sarah take apricot in chimps' symbols](image)

The shapes are relatively easy to learn for several reasons. The number of shapes to be learned is small, about 130. These shapes are easily discriminated, not only because their quantity is small but also they vary in three visual dimensions: shape, color, and size. Each shape as a whole pattern, without any kind of analysis of its internal structure, is paired with its meaning. The meanings of words tend to be concrete and simple. The words need not be pronounced. Finally, to "write" a sentence, required shapes are simply arranged in a sequence, as in Fig. 1.

Blissymbols for the cerebral-palsied

A system called Blissymbols was invented by Bliss (1965) as a universal written language. Some symbols are simple outline drawings of objects (Fig. 2a); some are compounds of two or more simple symbols (b); some are arbitrary symbols for unpicturable and/or abstract concepts (c); and some are compounds of two or more of these simple symbols (d).

![Figure 2 — Types of Blissymbols](image)

The system of Blissymbols, like several other "universal languages" invented over the
past few hundred years, has never caught the public’s fancy. However, a limited number of Blissymbols (up to 500), are used by some cerebral-palsied children who cannot articulate speech (Kates, MacNaughton, & Silverman, 1978). A child in a wheelchair is provided with a table of symbols in a tray that sits on his or her lap. To communicate a message (e.g., “I want water”), the child simply points to required symbols in sequence. Underneath each symbol is the English word written small to help English readers communicate with the symbol user. The words could, of course, be written in any language.

Learning 200 or so Blissymbols is not difficult, because each simple shape need not be analyzed or be associated with sounds. Learning complex Blissymbols, especially those of types (c) and (d), might be difficult for some handicapped children of low IQ. Several mirror-image Blissymbols represent two concepts that are similar save one subtle aspect. The case in point is a left slanted line and a right slanted one representing the and a (or is it other way around?) respectively. Some Blissymbols, such as Fig. 2d, are formed by absurdly elaborate embedding, requiring analysis and time to process. Without analysis, such symbols can be confused with other symbols sharing a few of the same embedded elements.

The Blissymbols do not appear to be the optimal means of communication for severely language handicapped children. (For “Optimal script for the language-handicapped,” see Taylor, in press.)

**Chinese Characters**

The only full logography used in modern world appears to be the Chinese system. It has been used continuously for over 4,000 years. Today it is the sole writing system in the Republic of China and Taiwan, the major system in Japan, and a supplementary one in South Korea.

**Chinese characters in Chinese, Japanese, and Korean**

By way of preview, Fig. 3 shows the same sentence in Chinese, Japanese, and Korean.

a 我是中国人。

b 私は中国人です。

c 나는 중국인이다.

d 나는 중국인이다.

Figure 3 — *I am a Chinese* in Chinese (a), Japanese (b), and Korean (c, d)

Note that the word *Chinese* is identical in the three languages, though pronounced somewhat differently: /zhong guo ren/ in Mandarin, /chu goku juin/ in Japanese, and /chung guk juin/ in Korean. Only Chinese has tone variations, between four and nine, depending on dialect.

Each Chinese character (henceforth, character) represents one morpheme. But one Chinese word (or Chinese loan word in Japanese and Korean) is not always one character: It often consists of 2-3 morpheme characters. For example, the word ‘Chinese’ requires three characters (Fig. 3), each of which is one morpheme (‘middle, kingdom,
man'). The first two characters alone mean 'China'. These two plus the character for 'language' means 'Chinese (language)'.

One character represent one spoken syllable in Chinese and Korean, but can represent more than one in Japanese. Compare the character for 'kingdom, nation' in Fig. 3: It is /guo/ (CV) in Chinese, /guk/ (CVC) in Korean, but /goku/ (CVCV) in Japanese. (C=consonant, V=vowel)

The Japanese and Korean sentences require postpositions that indicate case roles of nouns, and verb endings that inflect for tense, levels of politeness, and so on. The sentences in Fig. 3 end in present tense verbs in two alternate levels of politeness, a neutral level before the slash (/), and a polite level after it. A writer chooses one level or the other, depending on his or her relation with the addressee. There are a few more levels of politeness, as indicated by a series of dots. The Chinese sentence requires neither postpositions nor verb endings that inflect.

The "grammatical morphemes" (inflectional endings, postpositions, function words, etc.) in Japanese and Korean tend to be written in phonetic scripts. Indeed, it is the necessity of expressing grammatical morphemes that compelled Japanese to invent phonetic scripts. Japanese, but not Koreans, tend to use Chinese characters even for native words, such as 'I' in Fig. 3.

Here we have seen how three languages can use the same script, similarly for some purposes, and dissimilarly for others. The different uses are necessitated by differences in languages: the Chinese language belongs to the Sino-Tibetan Family whereas Japanese and Korean belong to the Ural-Altaic Family. Japanese and Korean are similar only in syntax, but not in speech sounds, and hence not in vocabularies. Speakers of the three languages cannot communicate to each other by speaking; they can communicate to a limited degree through writing in characters. I now consider in more depth the use of Chinese characters in these three languages.

**Chinese characters in China**

A pure logographic writing system is feasible for the Chinese language, which does not require many grammatical morphemes, in particular inflectional endings. It suits the language, whose monosyllabic morphemes can be represented by individual characters. The logographic system is also useful for Chinese people in two ways: Speakers of mutually unintelligible dialects can communicate through writing, and many homophones can be distinguished when written. Because the Chinese language uses only about 400 syllable types, mostly V or CV, it has many homophones. For example, there are as many as 188 morphemes with the sound of /yi/. Even when one considers only common morphemes and differentiates them through tone variation, there are still 30 yi (C-M. Cheng, 1982).

When each character represents a different morpheme, there must be as many characters as there are morphemes in a language. The number of characters is estimated to be about 50,000. For daily use, however, about 3,000, and for scholarly use, about 6,000, may be required.

Note that the number 3,000 refers to characters, not to words. Because a word often consists of 2-3 characters, and each character can appear in a number of different words, to know 3,000 characters implies that one knows many times that number of words, as shown in Fig. 4. In a count in Taiwan, 40,032 words were found in regular
use, but they used only 4,532 characters (Liu, Chuang, & Wang, 1975).

If several thousand character are to be visually discriminated, many of them have to be complex. But some highly complex ones seem to have been created to represent esoteric meanings, and hence are needed infrequently, if at all.

In the Republic of China (but not in Taiwan), as part of reforming the writing system, over 2,000 common characters have been simplified, from an average 16 strokes to 10 (C.-C. Cheng, 1977). Apparently, some individuals simplify characters at whim, causing much confusion. Some drastically simplified characters, whether standard or nonstandard, are no longer recognizable to overseas Chinese, Japanese and Koreans. Table 2 shows a sample of one judiciously, and five injudiciously, simplified characters. The informative “innards” have been removed from the latter, rendering them not only indiscriminable from each other but also hollow and unbalanced.

Table 2. Original and Simplified Characters

<table>
<thead>
<tr>
<th>Original</th>
<th>Simplified</th>
<th>Strokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

(Taylor, 1981, p. 11, with permission of Academic Press)

Since a character is recognized as a whole pattern, a simplified character is not necessarily easier to recognize than a complex one; in fact it is harder to recognize because it contains fewer cues for discrimination than does a complex one (Fukuzawa, 1968; Kawai, 1966). Simplification certainly reduces the time for writing characters. But we are entering an electronic age: Even Chinese characters can be typed on a word processor at a rate that does not depend on their complexity, though not as fast as phonetic symbols (Hecker, 1984; see below).

Characters appear hopelessly complex, numerous, and arbitrary to those who do not use them. Actually, characters are formed according to six principles or categories, which
Impart some system. Table 3 lists the six categories along with examples.

Table 3. Six Categories of Characters

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>pictograph</td>
<td>☀ 日 sun</td>
</tr>
<tr>
<td>Simple Ideograph</td>
<td>日 上 above</td>
</tr>
<tr>
<td>Compound Ideograph</td>
<td>日, 日 → 明 bright (sun, moon)</td>
</tr>
<tr>
<td>Analogous or Derived</td>
<td>网 fish net → network → cobweb</td>
</tr>
<tr>
<td>Phonetic Loan</td>
<td>米 /mai/ wheat</td>
</tr>
<tr>
<td>Semantic-Phonetic Compound</td>
<td>女, 女 → 妇 woman /ma/</td>
</tr>
</tbody>
</table>

Here four important categories are described.

1. Pictographs are iconic representations of concrete objects such as the sun and the moon. In some characters, the iconic origins are still discernible, in others less so, and in a great many others, not at all. Fewer than 3% of all characters are pictographs (Tsien, 1962).

2. Simple ideographs express relational or abstract concepts (e.g., 'above' and 'below') that cannot be easily depicted by pictures.

3. Compound ideographs contain two to four ideographs or pictographs. Thus, two pictographs, one for 'sun' and one for 'moon', join to form the character for 'bright'.

4. Semantic and phonetic compounds form the majority (80–90%). A sound-cuing phonetic (/ma/) and a meaning-conveying radical ('woman') are joined to form the character 'mother' pronounced /ma/.

These principles do not enable a reader to discover the meaning and sound of a new character, but they do serve as mnemonics for remembering already learned characters.

Learning characters in China

To make learning manageable, one must first designate a limited number of common characters, about 3,000, as 'educational' to be taught in primary schools. The educational characters are those used in popular newspapers and other publications. Taught in secondary schools are 3,000 or so additional characters, which tend to appear in scholarly publications.

Learning a few hundred logographs is easy; even chimps can learn them (Premack & Premack, 1972). But learning 3,000–6,000 characters takes time and effort. According to the director of the Language Research Institute in China, about 30% of school time is spent mastering characters (cited in Ohara, 1980). Each shape must be discriminated
from a few thousand others; each shape is associated not only with meaning but also with sound; it must be analyzed into semantic and phonetic components, as well as into a series of strokes in writing.

Recognition of individual characters can be learned by the Three-Phased Learning method: whole → analysis into components → mature whole (Taylor & Taylor, chap. 10). Initially, a character as a whole visual pattern is associated with its sound (syllable) and meaning (morpheme). In the analysis phase, ideographic compounds and semantic-phonetic compounds are analyzed into their components. Each character can be further analyzed into an ordered set of strokes in writing. In the mature whole phase, the components come together again in wholistic perception, based on secure knowledge of the components and strokes. Whereas the initial global shape perception can differentiate only a small number of classes of characters, the mature global perception based on components can make much finer distinctions, permitting a large vocabulary to be recognized.

Because character writing involves analysis of a pattern into an ordered sequence of strokes, as well as full recall of patterns and fine motor coordination, it requires much practice. Accordingly, character writing is regularly assigned as homework.

In not too distant a future, schoolchildren may learn to write on a word processor, thus lightening their heavy homework load. For some Chinese children in North America, the use of word processors to learn to write is in fact an event that is taking place now. All the same, learning how to write characters should never be abandoned, for it helps in learning to recognize characters. It is part of the analysis phase in the Three-Phased Learning. In Liu's (1978) experiment, fourth graders who learned characters in three aspects (sound, meaning, and writing) scored higher on tests of sound and meaning than those children who learned, in the same interval of time, only sound and meaning.

Because characters do not code sounds directly, some kinds of phonetic scripts are required for teaching the sounds of characters. (The phonetics in semantic-phonetic compounds are unreliable clues to the sounds.) Two kinds of auxiliary phonetic scripts are used today: In the Republic of China, “pinyin” (the letters of the Roman alphabet), and in Taiwan “zhuyin fuhao” (37 symbols that represent the initial C and the final V or VC). These auxiliary phonetic scripts are learned before the characters, and then are used to annotate the characters to be learned.

Pinyin can be used also in word processing. A character or word is typed using pinyin, and a special “lookup” key is pressed to display, on a “virtual keyboard,” several common characters or words that have the same sound as the target. The writer chooses the one that fits the context of a text. If those several displayed do not include the target, the writer presses the special key for a new virtual keyboard displaying less common alternatives (Becker, 1984). So character entry is a two- or three-stage operation, to compare with the one-stage operation for pure phonetic scripts. The large number of characters, and the abundance of homophones necessitate this multi-stage operation.

Complex as is the Chinese system, it does not prevent a nation from attaining full literacy. In Taiwan, the illiteracy rate is a low 0.4%; it used to be 20% in 1950 (Ministry of Education, cited in Liu, 1979). In the Republic of China, the illiteracy rate was 23.5% at the last mammoth census of 1982 (reported in The Globe and Mail, 1982, Oct. 28); it is estimated to have been 80% before the birth of the Republic of China. (An illiterate is a person, over 12 years of age, who cannot read or can read
only a few characters.) Obviously, raising a literacy rate depends a great deal on improvement of political, social, economic conditions.

**Kanji in Japan and Korea**

Chinese characters were adopted by Koreans in the 1st century A.D., and were introduced to Japan by Koreans in the 5th century A.D. Today characters form the major writing system in Japan, and a supplementary one in South Korea. Chinese characters are called "Kanji" in Japan, and "Hancha" in Korea; both terms mean 'letters of Kan/Han dynasty'. I will use Kanji, as it is the better known of the two terms. The term Kanji refers to the script as well as to individual characters, single or plural.

Kanji learning in the three countries is basically similar, in that each Kanji is learned by the Three-Phased Learning, along with the principles or categories of Kanji formation (Table 3). There are some dissimilarities, however. In Japan the number of "official" Kanji is about 2,000, half of which are learned in primary school, and the remaining half in middle school. Educated people end up learning additional "unofficial" 1,000 Kanji.

Japan has bona fide phonetic scripts, called Kana (see below), which are used to annotate the sounds of Kanji to be learned. The sounds of Kanji are harder to learn in Japan than in either China or Korea: The majority of Kanji have at least two different sounds, one Chinese-derived and one Japanese-native. Chinese-derived sounds approximate the Chinese sounds of Kanji, whereas Japanese-native sounds are none other than Japanese words for the concepts expressed in Kanji. And many Kanji have few variants of Chinese sounds as well as of Japanese sounds (Taylor & Taylor, 1983; chap. 4).

Take the three Kanji that together make up the word Chinese in Fig. 3. They are pronounced in the Chinese way as /chu goku jin/. In variants of the Chinese sounds, /jin/ is pronounced as /nin/ and /goku/ as /kok/ in words such as /nin pu/ ('man worker' = 'coolie') and /kok ka/ ('national anthem'). Then, each of the three Kanji for 'middle kingdom man' has Japanese sounds: /naka, kun, hito/. /Hito/ itself becomes /bito/ in /hito bito/ ('man man' = people). To complicate the picture further, a reader is not always sure when a Kanji is to be read in a Chinese or a Japanese way. When a word is made up of more than one Kanji, each may sometimes be read by a different method.

Learning 2–3,000 Kanji, each of which has multiple sounds, takes time and effort. And mastery is seldom complete: Even highly educated people may not know unusual sounds of some Kanji, or cannot fully recall infrequent Kanji for writing. Word processing on a computer is similar to that described for Chinese, except that in Japanese, grammatical morphemes narrow down alternatives to a target word (Becker, 1984; "Learning characters in China").

However, learning by whole pattern a few hundreds of Kanji for simple concepts is easy, especially if the sounds are bypassed. Even language-handicapped toddlers and preschoolers can learn them (see "Optimal script for the language-handicapped," Taylor, in press) In 10 months, three normal toddlers (aged 1;6, 1;8, 2;5 [2 years 5 months]), tutored by their mothers under Steinberg's guidance, learned 300 500 Kanji and Kana words (Steinberg, Yoshida, & Yagl, 1985). Some preschoolers "pick up" up to 170 Kanji before entering school (Muralshl & Amano, 1972; see "Kana" below).

In South Korea, Kanji learning is relatively painless, because the official Kanji, 1,800, are sufficient, and because each Kanji has only one Chinese-derived sound. The official Kanji are taught in middle and high schools, not in primary schools. Even official
Kanji do not appear in government publications and books for children; they appear in scholarly publications and the political and economics sections of newspapers.

In Japan and Korea, there have been sporadic movements to reduce drastically, or even abolish, Kanji. Kanji are in fact abolished in North Korea. Actually, abolishing Kanji may not be a good idea. The limited number of Kanji are not too difficult to learn; they also serve useful purposes, such as disambiguating homophonic Chinese loan words or making key content words stand out in texts (see Fig. 3 above). They also provide "the magic key" to the written treasures of the past, and serve as a means of communication among speakers of three different languages.

**Syllabaries**

In a syllabary, one symbol represents one syllable. Languages differ enormously in the number of syllables used, from as few as 100 (e.g., Japanese) to as many as a few thousand (e.g., English, Korean). A pure syllabary is feasible only for a language with a limited number of syllables such as Japanese. Since the number of symbols required is relatively small, the shapes of the symbols need not be complex.

Of the several varieties of syllabaries used today, I will discuss five: Cree, Eskimo, and Val briefly, and two Japanese Kana at some length. A sample of CV symbols in these syllabaries is shown in Table 4 in Vowel (row) x Consonant (column) matrix or chart.

<table>
<thead>
<tr>
<th>Table 4. Symbols for CVs in Five Syllabaries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Cree</td>
</tr>
<tr>
<td>Eskimo</td>
</tr>
<tr>
<td>Val</td>
</tr>
<tr>
<td>Hiragana</td>
</tr>
<tr>
<td>Katakana</td>
</tr>
</tbody>
</table>

For the Val syllabary, see Scribner & Cole (1981, Fig. 3.2); for Cree and Eskimo, see Jensen (1970, Fig. 203, 206); for Kana, see Taylor & Taylor (1983; chap. 4).

**Cree-Eskimo syllabary**

The Cree-Eskimo syllabary was invented in the 1840s by Rev. J. Evans for the Cree Indians and their neighboring tribes. The syllabary has 44 simple geometric shapes for 44 open syllables, as shown in Table 4. In 1885 the Rev. Peck modified the system (using 48 symbols) for the Eskimos of the eastern Arctic. In both scripts, marks have to be attached to the simple shapes to represent some closed syllables (CVCs).

Each CV symbol can be analyzed into C (coded by the shape) and V (by the orientation of the shape). Learning simple shapes distinguished by left-right or up-down orientation may appear easy, but it actually confuses young children (Gibson, Gibson, Pick, & Osser, 1962; Tanaka & Yasufuku, 1975). In reproducing 12 Cree CV symbols (Taylor & Taylor, 1983, Table 2-2), I myself made an orientation confusion error by listing for /ka, ke, ki, ko/ 'p q b d', which should have been (I believe) 'p q b d'. Oops, wrong again! They should be 'b q p d': I checked and re-checked Jensen's (1970) Fig. 203. It is even more confusing when the symbols for /na ne ni no/ are the same 'b q p d' but lying on their sides!
Vai syllabary

The Val people in Liberia, Africa, have a syllabary that evolved from a picture-word writing. The invention of the syllabary in the early part of the 19th century is credited to a native Val (Gelb, 1963; Jensen, 1970).

The syllabary consists of approximately 210 symbols for open syllables (Scribner & Cole, 1981, Fig. 3.2). As can be seen in Table 4, each symbol as a whole pattern codes a CV syllable; that is, each symbol is not analyzable into C and V components, as in the Cree-Eskimo syllabary. With a few exceptions, symbols for related sounds are not similarly shaped.

Apparently Val speakers require 2 to 3 months of lessons to achieve some functional literacy in the script (Scribner & Cole, 1981). In spite of the simplicity of the syllabary, literacy in the Val script is low, and is confined largely to two thirds of male adults. The Val script is not the tool of formal education; nor is it the tool of obtaining knowledge from printed texts. In other words, Val-script literacy is not "the magic key that unlocks the door to the wonderland of stories and information" (Taylor & Taylor, 1983: 397). Instead, it is used for such mundane activities as letter writing and record keeping. Val-script literates tend to be farmers who also do crafts or grow cash crops. Literacy is perhaps a unnecessary luxury for most Val people who are subsistence farmers.

Kana

For a few hundred years after the adoption of Kanji, Japanese scholars read and wrote using Kanji exclusively. But the practice was extremely awkward, because Kanji are not suited for representing grammatical morphemes (see Fig. 3). Thus the Japanese created two syllabaries between the 8th and 12th century A.D.

The Japanese syllabary, called Kana (from kari na, 'false name' or 'borrowed name'), comes in two variants, Hiragana and Katakana (Table 4). (The term Kana is used for the scripts as well as for individual symbols, singular or plural.) Each Hiragana symbol is a copy of a cursive Chinese character, whereas each Katakana is a side or fragment of a character. The two variants of Kana differ in the shapes of symbols and also in use, as we shall see. Otherwise, there is a one-to-one correspondence between the symbols of the two Kana sets.

Kana symbols code syllables, all open, except N. There are 46 basic Kana; to 25 of them marks can be added to create secondary Kana; several of the basic Kana can be written small to represent assimilated (double Cs) and contracted sounds, as in ge.k.ki.yu (Kana for the first 'k' is 'tsu' written small). In all there are 108 symbols in each Kana set.

Kana is extremely easy to learn for five reasons:

- Its number of symbols is small.
- The shapes of symbols are simple yet discriminable.
- Its symbol codes a syllable, a stable phonetic unit that does not change its sound value in different phonetic contexts.
The syllables coded are all open (except N).
- Symbol-syllable correspondence is nearly perfect: each symbol consistently codes the same syllable, and each syllable is consistently written by the same symbol.

Kana is so easy that many children learn it, or rather "pick it up," before entering school. In a large survey of preschoolers' reading activities, close to 90% of the children, a month before entering school, could read 60 or more Hiragana (as well as 8-168 Kanji) (Muraishi & Amano, 1972). Armed with the knowledge of 60 or more Hiragana, the children should be able to negotiate simple stories written in that script.

In the above survey, the writing system was not the only factor that influenced the preschoolers' reading achievement. A few other factors that mattered were: gender (in favor of the girls over the boys); the fathers' education (in favor of the higher over lower education); the years spent at a nursery school (in favor of 2-3 years over 1 year). About the last factor, the children are not taught reading at the nursery schools; rather, they are exposed to printed words and sentences that identify the owners of objects and give simple instructions. The children's questions about the printed materials are also answered. The moral: When printed materials are used to convey vital information, preschoolers are motivated to learn them.

Hiragana are used mainly to represent grammatical morphemes while Kanji represent key content words (see Fig. 3). Katakana are used to represent European loan words and onomatopoetic words. This kind of mixed text helps a reader allot his or her attention over text differentially, more to key content words (visually complex Kanji) and less to grammatical morphemes (visually simple Hiragana).

Japan boasts one of the highest literacy rates in the world, 99%, with illiteracy confined largely to the mentally retarded (Sakamoto & Makita, 1973). One factor responsible for this salutary event could be initiation into reading with the simple Kana; another factor must be the high socio-economic status of most Japanese people.

**Alphabet**

**Alphabet: its nature and variety**

In an alphabet, each letter codes one phoneme, in principle but not necessarily in practice. Since most languages of the world use between 20 and 37 phonemes (Maddison, 1984), an alphabet needs only a small number of letters. These letters can therefore be shaped simply.

Today the alphabet has all but conquered the world: It is used in almost all Indo-European languages as well as in such non-Indo-European languages as Hebrew, Finnish, Turkish, and Vietnamese. The widespread adoption of an alphabet does not necessarily attest to its superiority over other writing systems. Most languages require grammatical morphemes, and hence a logography is impractical. This inconvenience is not the only reason why a logography is impractical; far from it (see "Logography"). Most languages tend to have over 1,000 syllable types, and hence a syllabary is also impractical. Thus they have no choice but adopt an alphabet.

There are a variety of alphabets: some differ in letter shapes (Roman, Cyrillic, Devanagari, Arabic, Hebrew, etc.); in most alphabets, C and V letters retain their separate shapes in syllables, but in a few (Devanagari, Tamil), C and V letters fuse to form CV composites; most have letters for both consonants and vowels, but a few
(Hebrew, Arabic) have letters primarily for consonants. Vowel-less alphabet and orthography are possible for these Semitic and Hamitic languages, in which Vs indicate inflection while Cs indicate root morphemes.

All these alphabets can be traced back to the Semitic syllabary–alphabet of 1,000 B.C.

**Learning to read in an alphabet**

Learning an alphabet itself is easy, because it has a small number of simply shaped letters. Many preschoolers learn it before entering school. Learning to read in an alphabet is not necessarily easy for several reasons.

In some alphabets such as Hebrew and Arabic, Vs are indicated not by proper letters but by optional marks over, under, or inside Cs. Because of the "invisibility" of the tiny marks "tucked under" C letters, beginning Hebrew readers make many errors in sounding out Vs (Feitelson, 1980). (For adults, however, the presence or absence of the V marks does not affect the latency for pronunciation; Navon & Shimron, 1981).

The phonetic unit coded in an alphabet, the phoneme, is abstract, small, and unstable. Stop Cs by themselves are unpronounceable, and most phonemes change their sound values in different phonetic contexts. As part of the phonics method of reading instruction, phonemes must be put together into syllables and words by a process called phonetic blending. Preschoolers find it easier to blend syllables into words (e.g., *candy*; *teacher*) than phonemes (e.g., *pay*; *up*) (Brown, 1971; Coleman, 1970). Finally, several meaningless phonemes must be strung together before the meaning of a word emerges.

Understandably, the phoneme is not the unit people most easily become aware of. Historically, the alphabet was the last writing system to emerge, after logography and then syllabary (Gelb, 1963; Jensen, 1970). Children find it more difficult to segment a word into phonemes than into syllables (Liberman, Shankwell, Fischer, & Carter, 1974).

A single word in an alphabet tends to require a long array of letters, longer than a word in a syllabary or logography. For example, *gentleman* requires 9 letters in English, 6 Kana in Japanese, 3 syllable-blocks in Korean "Hangul" (see below), and 2 Kanji (in translation), as shown in Fig. 5.

```
<table>
<thead>
<tr>
<th>English</th>
<th>gentleman</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kana</td>
<td>セントルマン</td>
</tr>
<tr>
<td>Hangul</td>
<td>지엔블맨</td>
</tr>
<tr>
<td>Kanji</td>
<td>神士</td>
</tr>
</tbody>
</table>
```

**Figure 5** — *gentleman* in English, Kana, Hangul, and Kanji, in that order

The longer an array, the more material must be processed visually, posing a problem in sequencing and grouping letters and sounds for word identification. People often have trouble seeing and remembering the order of items in a list, especially in the middle. In English, transposition errors involving letters in medial positions, such as 'er-' and 're-' in *there* and *three*, are intractable among Grade 2 poor readers (Park, 1978 1979).

For an alphabet to be a useful representation of speech, its letters have to code sounds consistently and accurately. But preserving close letter sound correspondence is difficult in the face of sound changes that can occur over time, over geographic regions, and in
word derivation. For example, for heal and health, if the latter is spelled "helth" to represent its sound more closely, the relatedness of the two words becomes less transparent. The same observation applies to train, which is pronounced as "trine," which rhymes with vine, in Australia, and as "trane," which rhymes with rain, in most other English-speaking countries.

The ideal is to have perfect letter-sound correspondence in a language whose sounds do not change unless root meaning also changes.

Let us contrast two Roman alphabets, Finnish (an example of a high letter-sound correspondence), and English (an example of a low letter-sound correspondence).

**Finnish vs. English**

The Finnish alphabet has 37 letters (a few of which may be digraphs) for Cs, Vs, and diphthongs, plus 6 letters for foreign sounds. It has a high letter-sound correspondence, perhaps because writing was cultivated relatively late in the 16th century, and spelling was regularized in the 19th century (Kyöstio, 1980). About 15% of children already read at a Grade 2 level when they enter school at age 7. By the end of Grade 1, children can decode nearly any word, whether familiar, unfamiliar, or nonsense. Regular letter-sound correspondence is even more beneficial for spelling.

By contrast, the English alphabet and orthography have a dismally low letter-sound correspondence, for several reasons (Taylor & Taylor, 1983; chap. 6). To cite two reasons, the alphabet has only 23 letters to represent 44 phonemes (of the 26 letters, 'x, q, c' can represent or include the same phoneme as does 'k'). The sounds, especially vowels, changed substantially about 500 years ago while spelling remained more or less unchanged.

Today, each C letter represents on average 2.4 sounds, and each V letter 8.2 sounds (Dewey, 1971). See how 'a' changes its sound in about, farm, fat, fall, face, fare, hurrah, feat, instead, boat. This kind of letter-to-sound inconsistency causes difficulties in pronouncing. Even college students make more errors and take longer to read aloud irregular words (vise, should) than regular words of the same length and frequency of occurrence (glue, chant) (Baron & Strawson, 1976).

Each C sound has on average 9.1 spellings, and each V sound 10.7 spellings (Dewey, 1971). This kind of sound-to-letter inconsistency causes spelling difficulty. If the sound-letter relation were regular, schoolchildren and adults should be able to spell any word, even pseudowords that are pronounceable though meaningless. When children (Grades 2–5) and college students were asked to spell pseudowords to dictation, no group of subjects scored perfectly on all types of words. Regularly spelled CVC and CVCe words (fat, jate) were easy, but irregularly spelled words (jation, casical, cazieize) were difficult even for college students (Marsh, Friedman, Welch, & Desberg, 1980). These kinds of words must be spelled by analogy to known words, such as elation for jation.

Partly because of the difficulty of English orthography, "early readers" (children who read before going to school) are few, on the one hand, and "non-readers" (schoolchildren who cannot decode words) are many, on the other. Only 1 3.5% of children can read at Grade 2 level when they enter school at age 6 (Durkin, 1966). Once in school, up to Grade 3 4 most children are learning to decode words (Cafare, Venezky, & Chapman, 1969).

As for non-readers, one study reports that one-fifth of all Grade 3 children tested in
three schools in a racially mixed, urban inner city were "total nonreaders" (Gottesman, Croen, & Rotkin, 1982). Another study reports high rates of non-readers among black working-class children in Philadelphia (Baron & Treiman, 1980). The non-readers could not decode any of the test words, which were familiar exception words (was, said, come) regular words (has, maid, dome), and pseudowords (mas, haid, gome). (All the studies on English orthography cited above have been carried out in the United States.)

The irregular and complex letter-sound relation of an alphabet does adversely affect learning to read and spell, especially in unfavorable socio-economic conditions.

**Alphabetic Syllabary**

**Hangul syllable-blocks**

In an alphabetic syllabary, each symbol codes a phoneme, as in an alphabet, but between two and four alphabetic symbols are packaged into a syllable-block, which is the unit of printing and reading. The Korean script called Hangul appears to be the only alphabetic syllabary. It was created in the middle of the 15th century A.D. by King Sejong, taking into consideration the articulatory features of the Korean sounds and Oriental philosophy (Taylor & Taylor, chap. 5). And it was created with the noble sentiment of helping ordinary citizens acquire literacy.

Table 5 shows a partial C x V chart or matrix.

**Table 5. Part of Hangul CV Syllable-Block Chart**
(with 3 of the 19 Cs and 4 of the 21 Vs)

<table>
<thead>
<tr>
<th>C</th>
<th>a</th>
<th>ya</th>
<th>o</th>
<th>yo</th>
</tr>
</thead>
<tbody>
<tr>
<td>ㅏ</td>
<td>ㅏ</td>
<td>ㅓ</td>
<td>ㅣ</td>
<td>ㅣ</td>
</tr>
<tr>
<td>ㅑ</td>
<td>ㅑ</td>
<td>ㅐ</td>
<td>ㅔ</td>
<td>ㅣ</td>
</tr>
<tr>
<td>ㅓ</td>
<td>ㅗ</td>
<td>ㅔ</td>
<td>ㅣ</td>
<td>ㅣ</td>
</tr>
<tr>
<td>ㅗ</td>
<td>ㅏ</td>
<td>ㅣ</td>
<td>ㅣ</td>
<td>ㅣ</td>
</tr>
</tbody>
</table>

Unlike the five syllabaries (Cree, Eskimo, Vai, Katakana, and Hiragana) shown in Table 4, in the Korean alphabetic syllabary, the two margins of the chart can show Hangul alphabetic symbols, Cs vertically and Vs horizontally. (Some of the V symbols are themselves combinations of two or more simple V symbols.) Each of the 19 C symbols can combine with each of the 21 V symbols to produce 399 (19 x 21) CV syllable-blocks. Then, underneath each of these CV blocks the final C can be placed to produce 7,581 (399 x 19) CVC blocks, only about one third of which are actually used. Underneath a handful of CVs, a second C can be placed to produce CVCC blocks.

Table 6 shows four alphabetic symbols that are packaged into syllable-blocks of three levels of phonetic complexity (CV, CVC, CVCC), with commensurate levels of visual complexity. With increasing levels of complexity, the likelihood of one syllable-block by itself unambiguously representing a morpheme increases.
Table 6. Hangul Syllable-Blocks in Three Complexity Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Alphabet symbol</th>
<th>Syllable Block</th>
<th>Syllable Structure</th>
<th>Morpheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>t</td>
<td>OT</td>
<td>V /a/ suffix; ah</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>Ce</td>
<td>C1</td>
<td>CV /da/ e11</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>C2</td>
<td>C1</td>
<td>CV C /Dal/ moon</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C1 C2</td>
<td>C1</td>
<td>CVCC /Dal9/ hen</td>
<td></td>
</tr>
</tbody>
</table>

In the original table (Taylor, 1980: 70), there was a typographical error: the Hangul for C2 should be as in this table, not C2. In this syllable-block, the empty circle is required to show that a vowel is alone in a syllable.

The three levels of complexity helps word recognition. In an experiment, a target syllable-block is recognized faster against the background of other blocks in all three complexity levels than in only the complexity level of the target (Taylor, 1980).

**Learning and using Hangul**

Hangul is easy to learn for reasons such as:

- Symbol–sound correspondence is high.
- The reading unit is a concrete and stable syllable.
- Although the number of syllable-blocks is large, each syllable-block need not be learned by rote; it can be deduced from the chart shown in Table 5.

Because of the systematic placement of 2-4 alphabetic symbols in a syllable-block, word processing on a computer is easy and fast (Chong, Han, & Kang, 1983). A writer types words in alphabetic symbols, and a computer packages them into syllable-blocks.

Some texts in South Korea, and all texts in North Korea, are written in Hangul only, without Kanji (Fig. 3d). All-Hangul texts are not as hard to read as all-Kana texts for three reasons. First, the three varied levels of visual complexity of syllable-blocks make discrimination easy. Second, a word tends to be short in syllable-blocks (see Fig. 5). Third, in Hangul texts, an extra space is left between each phrase (one content word and its accompanying grammatical morphemes).

Korea enjoys a high rate of literacy, with illiteracy confined to the mentally retarded or to the very old who have not benefited from compulsory education. The simplicity of Hangul, as well as favorable socio-economic conditions, must be responsible for this salutary event.

**Conclusion and Announcement**

I have described 4 writing systems and 10 scripts, and developed the following three themes:
1. The characteristics of language dictate, to some degree, the type of script.

2. The way a script represents meaning and/or sound has a great influence on how word recognition in that script is learned.

3. However, a script is not the sole influence on reading achievement and literacy attainment; socio-economic factors too play a role.

In my seminar, I will discuss research carried out on some of these scripts.
References


Becker, J. D. Computerized typing and editing can now be extended to all living languages of the world. *Scientific American*, 1984, 251, 96-107.


Liu, I.-M. Personal communication, February 7, 1979.


Taylor, I. The psychology of scripts. In D. de Kerekhove, & C. Lumsden (Eds.), *Western literacy, brain and culture*. New York: Springer Verlag, In press.


