Writing systems

Introduction

In the modern industrial society, writing is so obviously important that we take it for granted. An illiterate person is viewed as seriously handicapped. Yet, until about 200 years ago, the majority of people were illiterate, and in some parts of the world this is still true. For most of us, however, modern daily life depends heavily on writing where it is central in education and in many types of work as well as in providing us with a significant source of pleasure. We are surrounded by newspapers, magazines, books, signs, and computer text.

We must distinguish carefully between writing and language. Language is an innate ability of human beings. We all learn to speak with no formal training. Writing, however, is not innate; it must always be consciously taught and learned. Children only learn to read and write some years after they have learned to speak.

Language is a complex system relating sound and meaning. Writing is a graphic representation of a linguistic utterance. This definition of writing rules out pictures as writing. Pictures or drawings may indeed communicate, perhaps reminding the viewer of a story or event, but they are not writing in this sense because they do not represent specific linguistic utterances. A picture of a man fishing might be read in English as *The man is fishing*, *The man hopes to catch a fish*, *The man enjoys fishing*, or many other ways. The sentence *The man is fishing*, however, can only be read aloud in one way; to read this sentence as *The man hopes to catch a fish* or even with such a small difference as *The man was fishing* would be regarded as wrong.

History

Writing is relatively recent in human life, no older than about 5,500 years; human beings were speaking millennia earlier. Writing has only been invented three times from scratch. Much more often it has been borrowed from and applied to a different language. The invention of writing requires acquiring the notion that symbols can represent linguistic units, e.g., words, and then ways must be developed for writing any word in the language. If writing did not exist today, we might possibly create a writing system for English in this way: for the word *eye*, we might create a picture of an eye (see Figure 1); such a creation is called a pictogram, an element of a writing system, because it is a graphic way of representing a specific linguistic utterance, namely the word *eye*.

Then, we might use semantic extension to extend the meaning of this symbol to other words of a similar meaning, such as *see* or *vision*. We might also extend the meaning of this symbol to the pronoun *I* using phonetic extension. In both types of extension, we would have to rely on the context to tell us which word was intended. If sorting out these different meanings for this symbol became too difficult, we might differentiate them with extra marks. For the verb *see*, we might add an arrow to indicate symbolically the action of a verb. For *I*, we might make a compound of the eye and a stick figure for a person. Using these and other devices, we could create symbols adequate to write an entire language. Today, we can see that Sumerian cuneiform and Chinese characters were created using these principles.
We are certain that writing was independently invented three times. First, in Mesopotamia by the Sumerians about 3300 BCE. Second, in China about 1500 BCE. And finally by the Mayans in Mesoamerica (southern Mexico and neighbouring areas) between 500 and 300 BCE.

Living in Mesopotamia, the Sumerians had small clay geometrically shaped objects called tokens for accounting purposes. They made bookkeeping records by pressing these objects into clay tablets, then they began to draw the image with a pointed stick, and finally used a triangular stylus to make wedge-shaped symbols (see Figure 2). Their writing is known as cuneiform. Using the principles mentioned above, the Sumerians created a writing system capable of writing any utterance in the language. The Akkadians, speaking a very different language, conquered the Sumerians in the second millennium BCE and adapted cuneiform writing to their own Semitic language Akkadian. Some symbols represented morphemes, and some represented sounds, generally consonant-vowel sequences. Although Akkadian writing was extremely complicated and required considerable schooling to master, it enjoyed enormous success with the last known text written in the first century CE.

The earliest known Chinese writing is the oracle-bone inscriptions; these are texts on bone or shell predicting future events. Chinese writing today is essentially a direct continuation of this early writing. Although the inventory of characters and the calligraphic style of writing has changed over the centuries, the structural principles of the writing system have remained very much the same. Chinese writing was borrowed by neighbouring cultures and adapted for Korean, Japanese, and Vietnamese.

The surviving documents from Mesoamerica in Maya are primarily stone tablets of an historical
nature; most texts were written 250–900 CE. These texts are notable for their very careful calendrical details. Circumstances in the Mayan world changed causing writing to become considerably less common around 1000 CE, and it died out entirely around 1600 although Mayan languages continue to be spoken in the area today. Knowledge of the writing system was lost, and modern decipherment of the Maya texts only dates from the 1950s.

The Egyptians likely borrowed the notion of writing from the Sumerians around 3000 BCE although some scholars argue that Egyptian writing is an independent invention. Although Egyptian writing is also a mixture of morphographic and phonographic writing (see below), it is pictorially and structurally quite different from cuneiform writing. It is also quite complex, but it lasted until around 450 CE.

Semitic-speaking peoples from the eastern end of the Mediterranean likely acquired writing from the Egyptians around 1500 BCE. They simplified the system considerably to under 30 symbols; these were used to represent only the consonants; such a system is known as an abjad (see the discussion of Arabic below). This Semitic writing spread to all the Semitic languages in the area including Phoenician, Ethiopian, Hebrew, Aramaic, and Arabic. It spread eastwards to other languages across Asia. Most likely it is the source of, or at least had a strong influence on Brahmi of India which is the ancestor of all the native scripts of India, Tibet, and most of Southeast Asia. Today, this Semitic writing is used for several languages, principally Arabic and Hebrew; it is written in lines running from right-to-left.

The Phoenicians, a Semitic people of the Syria-Lebanon area, brought their script to Greece. The Greeks adjusted the Semitic writing system slightly by adding vowels producing the first alphabet and changing the direction of writing to left-to-right. The Greek alphabet was borrowed and adapted for several languages: Coptic, Armenian, Georgian, Slavic, and Gothic. Significantly, it spread to Italy and was adapted to Latin, becoming the Roman alphabet which spread throughout western Europe. Since the Middle Ages, the Roman alphabet has become the most widely used script in the world.

Today, all writing in the world is derived from either the Chinese or the Semitic scripts or was created in a social context where such scripts were in use.

**Structure**

**General**

Writing can generally be divided into two broad categories: phonographic – where the symbols represent units of sound in a language, or morphographic – where the symbols are related to morphemes. Phonographic systems are by far the more common. Most phonographic systems are alphabetic where a symbol relates to a single phoneme; the Roman, Greek, and Cyrillic alphabets are examples of these. But some phonographic systems are moraic, such as Japanese kana, where each symbol represents a mora, i.e. a CV (consonant-vowel) or C (final consonant) sequence in the language.

Writing systems seem never to be pure. Chinese writing is strongly morphographic, but it has a considerable phonographic aspect as well. The use of Arabic numerals in alphabetic writing is morphographic. Further, the common English way of writing different morphemes which sound the same differently is morphographic in nature in that it distinguishes the morphemes from each other: too – two – to; ewe – you – yew – U(-turn).

**Chinese**

Although Chinese writing requires the writer and reader to learn a sizable number of characters, it has a rather simple relationship to the language. The rule is that each syllable is written with one character, and each character represents one syllable. The majority of Chinese morphemes have only one syllable and thus are written with one character. Although in Chinese it is very common for different morphemes to be homophonous, i.e. to sound the same, these different, but same-sounding morphemes are written with different characters.

Some morphemes have two syllables and thus are written with two characters. Words in Chinese commonly consist of more than one morpheme and are thus also written with more than one character.

Until 1900, Chinese was normally written in a dialect known as Classical Chinese; this dialect
was not mutually intelligible with any spoken variety of the language. Everyone spoke their native dialect, but wrote the same, i.e. in Classical Chinese. From around 1900, writing has been in the Mandarin dialect, essentially the dialect of the capital Beijing. Thus, for the speakers of Mandarin, who live in the north and west of China, they write much as they speak. For the speakers elsewhere, e.g., Shanghai or Canton, their spoken language is a different and mutually unintelligible dialect from their written Chinese. Today all Chinese is written in the same dialect – Mandarin; thus the unity of Chinese writing continues although the dialect used for writing has changed from Classical Chinese to Mandarin.

Chinese writing was borrowed by other cultures, and Chinese characters are still used strongly in Japanese writing, and to a reduced extent in Korean. Calligraphy is a highly valued art in Chinese, Japanese, and Korean societies.

Previously, Vietnamese was written with characters borrowed from Chinese together with many characters invented in Vietnam but not used elsewhere.

Since around 1900, Vietnamese has been written with a version of the Roman alphabet.

Arabic

Arabic is the most widely spoken Semitic language today. Like other Semitic writing, it is written in horizontal lines going from right to left. As a whole, the letters in a word must be joined and not written separately; this situation results in letters having various shapes depending on where they occur in a word. Structurally, the Arabic script is considered an abjad; an abjad is like an alphabet except that only the consonants are written. In Arabic long vowels are in fact regularly written, but short vowels are not normally indicated although there are dia- critics for showing the short vowels where desired. Short vowels are normally written in the Qur’an and in materials for children and learners. Calligraphy is highly valued in Arabic culture, and a large number of different calligraphic varieties exist. The expansion of Islam has meant that the Arabic script has been adapted for writing a large number of other languages, such as Persian, Urdu, and many others across Asia.

Arabic is diglossic in that speakers of Arabic mostly use their local dialect for speaking, but use a common dialect called Modern Standard Arabic for writing. The spoken and the written dialects are not mutually intelligible. Modern Standard Arabic is similar to, but not exactly the same as, the Arabic of the Qur’an. Modern Standard Arabic is used in university lectures, in the news on television, and in other relatively formal situations, but it would be felt to be pretentious for ordinary conversations. Similarly, it is perfectly possible to write a local dialect, but that would generally be regarded as undignified.
Devanagari

Devanagari is the script used for writing Sanskrit, Hindi, Marathi, and Nepali. It is typical of the scripts of south Asia. Structurally the script is called an abugida (see Figure 4). In an abugida, each consonant has a symbol. Vowels following a consonant are written with a diacritic on the consonant symbol. The short vowel /a/ is not written; thus the absence of a vowel symbol means that an /a/ is present. Initial vowels are written with special symbols. Consonant clusters are written by combining the consonant symbols into a ligature, i.e. a single symbol formed by combining two symbols. All the native writing systems of India are structurally abugidas.

asampatti ‘non-success’

<\a\ s\ m\ p\ i\ t\t>

Japanese

Japanese has the most complicated writing system in use today. Historically, Japanese borrowed Chinese writing. However, a character was typically borrowed along with the Chinese word it represented. A character was used to represent this Chinese word, but the same character was also used to represent the corresponding Japanese word. For example, the same character is used to write the native Japanese words /itsu/ meaning ‘five’, but in other contexts it is also used to write the borrowed Chinese word /go/ also meaning ‘five’ (see Figure 5). As a result of this history of borrowing, almost all Chinese characters have a Japanese reading (called the kun-reading) as well as a Chinese reading (called the on-reading). Today, the native Japanese word and the borrowed Chinese word are usually found in
different environments, although with the same meaning and they are written with the same character.

Over time, simplified forms of characters were used purely for phonographic purposes. This system known as *kana* now has a symbol for each *mora* in Japanese, i.e. for each CV cluster or syllable-final –C in the language. In principle, any utterance could be written in *kana*, but in fact Japanese is written with a mixture of characters (called *kanji*) and *kana*; often the root of a word is written with *kanji* and suffixes are written with *kana*, but some words are written entirely in *kanji* or *kana*. As a further complexity, there are two equivalent systems of *kana*, known as *hiragana* and *katakana*. *Hiragana* is the more neutral form and is used for writing suffixes, but *katakana* is used for emphasis, for certain onomatopoetic words, for telegrams, and for borrowed words. After World War II, the government moved to limit the number of characters in use to 1945 although in practice more are actually used, especially in proper names. Despite being a very complicated writing system, literacy in Japan is essentially 100 per cent. Children typically start school, already having learned to write *hiragana* at home.

---

**Cree**

Cree is an Algonquian language spoken in northern Canada. The script is often called Cree Syllabics. It was created in the nineteenth century by John Evans, a Methodist missionary. The script is widely used by the Cree and has been adapted for use by other languages such as Inuktitut.

The Cree script is unique in its nature. CV sequences are written by a single symbol. The consonant is shown by the shape of the symbol, and the vowel is shown by the orientation, i.e. the rotation of the symbol.

---

**Transliteration and Romanisation**

Scholars find it useful to represent foreign writing in the script of their own language. For example, English-speaking scholars often convert other scripts to some version of the Roman alphabet, a process known as Romanisation. Standard Romanisation exist, for most non-Roman scripts, although there are often more than one Romanisation in use. For example, until the late twentieth century, the Wade-Giles system was most commonly used for Romanising Mandarin Chinese (e.g., Mao² Tse²-tung¹); since then the Pinyin system has become standard (Máo Zédōng). Mandarin has four distinctive

---

**Figure 5** Japanese. The phrases for *five people* and *five days* are given in characters (*kanji*), *hiragana*, and *katakana*. The same character for *five* is used in both phrases, but it is pronounced as /go/ or /itsu/. /go/ was originally borrowed from Chinese and is called the on reading; /itsu/ is the original native Japanese word and is called the kun reading. The context tells the reader which reading is appropriate. In the lines below, the *hiragana* and *katakana* transliterations for each item are given.

<table>
<thead>
<tr>
<th>Characters (kanji)</th>
<th>Hiragana</th>
<th>Katakana</th>
</tr>
</thead>
<tbody>
<tr>
<td>五人</td>
<td>/go nin/ ‘five people’</td>
<td>ごにん</td>
</tr>
<tr>
<td>五日</td>
<td>/itsu ka/ ‘five days’</td>
<td>イツカ</td>
</tr>
</tbody>
</table>

---

Chinese

五人 五日 characters (kanji)

/go nin/ ‘five people’ /itsu ka/ ‘five days’

ごにん イツカ

한국어

사람 일

/son/ ‘five people’ /il/ ‘five days’

사리 일

日本語

人 曜

/gos/ ‘five people’ /itsu/ ‘five days’

ゴン イツカ
tones: these are indicated by small raised numbers in Wade-Giles, and by accent marks in Pinyin.

In certain situations, it is useful to distinguish transliteration, which shows the orthography, and transcription showing the pronunciation. Thus, the word for Japan in Japanese is /nippon/. This is written in hiragana with four symbols as にっぽん. Here, each hiragana symbol represents a single mora, or CV or –C sequence. The transliteration of this is <ni Q po N> . (Note that <Q> represents a syllable-final sound which is the same as the following consonant, and N shows a syllable-final nasal.) This transliteration provides a one-to-one relationship from each hiragana symbol to Roman letters; the spaces separate the morae. The romanisation, which shows the pronunciation, is /nippoŋ/. With this transliteration, there is a one-to-one relationship between the sounds of Japanese and the Roman symbols. Transliterations are useful in showing how the orthography works; transcriptions are useful in showing how the actual symbols work.

Romanisation is not the only type of change of script of course: speakers of languages using Greek, Cyrillic, Arabic, Hebrew scripts all convert foreign names to their language and script. Chinese assigns characters with a similar sound to represent foreign names (Figure 7).

Sociolinguistics

Many languages are diglossic, where typically one form of the language is used for writing and another for speaking. Chinese and Arabic have been mentioned above; others would be Swiss German and Tamil (south India and Sri Lanka). In German-speaking areas of Switzerland,
schools and universities are the only places where Standard German would be spoken, otherwise Swiss German is used; conversely, written Swiss German is unusual.

Languages always have dialect variation. Sometimes, one of these dialects is chosen, usually for social and political reasons, as the one conventionally to be used for use in writing. This dialect is often called the standard dialect. Old English (500–1100) writing was rather uniformly based on the Winchester dialect where the capital was located. On the other hand, Middle English (1100–1500) tended to reflect local usage with a wide range of dialectal variation with relatively little sense of a standard dialect. Towards the end of the middle English period, the dialect of the later capital London came to be used as the standard dialect. And after 1500, printing brought further uniformity to English spelling, using the London dialect of the sixteenth century, a usage which has continued to the present day. Despite a few dialectal variations in English spelling in the United States (e.g., colour – color; centre – center; defence – defense), English spelling is quite uniform around the world.

In the middle of the twentieth century, the People’s Republic of China (PRC) made significant reforms in the writing of Chinese by simplifying a large number of characters, substituting forms written with fewer strokes. These simplified characters are now the norm there. However, Taiwan viewed these characters as ‘communist’ and has continued to use the traditional forms of characters. People in Hong Kong as well as Chinese speakers outside China have continued using the traditional characters, although more from traditional rather than political motivation. Even the handover of Hong Kong to the PRC in 1997 has not yet radically changed the use of traditional characters there. Today, different computer codings exist for the traditional and simplified characters.

**Literacy**

Although many illiterate people lead happy and productive lives, being able to read and write is widely considered essential in the modern world. No clear answer exists to the question of what kind of writing system would maximise the efficiency of literacy education. Japanese has the most complicated writing system in use today, yet its illiteracy rate is close to zero. Spanish has a straightforward writing system with close to a one-to-one phoneme-symbol relationship, yet illiteracy has only recently been lowered significantly in many Spanish-speaking countries of Latin America. The choice of writing system is less important for literacy than the amount of money and time a society is willing to devote to teaching it.

H. R.

**Suggested further reading**
