3 Writing Systems

PETER T. DANIELS

Chapters on writing systems are very rare in surveys of linguistics – Trager (1974) and Mountford (1990) are the only ones that come to mind. For a century or so – since the realization that unwritten languages are as legitimate a field of study, and perhaps a more important one, than the world's handful of literary languages – writing systems were (rightly) seen as secondary to phonological systems and (wrongly) set aside as unworthy of study or at best irrelevant to spoken language. The one exception was I. J. Gelb's attempt (1952, reissued with additions and corrections 1963) to create a theory of writing informed by the linguistics of his time. Gelb said that what he wrote was meant to be the first word, not the last word, on the subject, but no successors appeared until after his death in 1985.[1] Although there have been few linguistic explorations of writing, a number of encyclopedic compilations have appeared, concerned largely with the historical development and diffusion of writing,[2] though various popularizations, both new and old, tend to be less than accurate (Daniels 2000). Daniels and Bright (1996; *The World's Writing Systems*: hereafter *WWS*) includes theoretical and historical materials but is primarily descriptive, providing for most contemporary and some earlier scripts information (not previously gathered together) on how they represent (the sounds of) the languages they record.

This chapter begins with a historical-descriptive survey of the world's writing systems, and elements of a theory of writing follow. Only one piece of theoretical machinery needs to be introduced in advance: the typology for categorizing the variety of scripts that have been used over the last five millennia or so. In the order they came into being, the six types of writing system are: *logosyllabary* (more precisely morphosyllabary), in which each character stands for a morpheme, and the characters can be used for the sound of the morpheme as well as for its meaning (in C. F. Hockett's formulation: "unit symbols represent syllables but with homophones distinguished" [1997: 381]) – there can be no purely logographic script; *syllabary*, in which each character stands for a syllable; *abjad* (the Semitic-type script), in which each character

stands for a consonant; *alphabet* (the Greek-type script), in which each character stands for a consonant or a vowel; *abugida* (the Sanskrit-type script), in which each character stands for a consonant accompanied by a particular vowel, usually /a/, and the other vowels (or no vowel) are indicated by consistent additions to the consonant symbols; and *featural* script (the Korean type), in which the shapes of characters correlate with phonetic features of the segments they designate.

Writing was independently invented at least three times, in West Asia, in East Asia, and in Central America. Details and references for the information summarized below can generally be found in *WWS*.[3]

1 Writing and History

1.1 Old world logosyllabaries and their relatives

The first known writing system was Mesopotamian cuneiform. The first language to be written was Sumerian. The first writing surface-cum-material was clay, and the first writing implement was a reed stylus of triangular cross section: a scribe would shape a suitably sized patty of clay and smooth its surfaces, then touch a corner of the stylus to the surface, leaving shallow wedge-shaped impressions (hence the name, from Latin *cuneus* 'wedge'). From one to a dozen or so wedges make up a single cuneiform sign. A limited repertoire of wedge orientations combine in a limited range of patterns that recur in the individual "signs" (but there is no connection between the patterns and the sounds or meanings represented by the signs: see figure 3.1).

The first recognizable documents come from about 3200 BCE from the city of Uruk, and the script remained in use, recognizably the same, down to at least the third century CE (Geller 1997). Each Sumerian sign (and there were something over a thousand of them) originally stood for a Sumerian word, and was a picture of the object named by the word. (It took a very short time – measured in decades – for the recognizable pictures, which were hard to draw with a stylus on clay, to turn into the patterns of wedges.) Signs for objects could also be used for related verbs: a leg could represent "walk," for instance.

	*xi*	"mix"	cf.	*ša*	
	'a		cf.	*an*	"god"
	*mud*	"fear"	cf.	*ig*	"doorleaf"
	*kun*	"tail"			
	*kam*	(number determinative)	cf. 	*be*	"if"
	*gil*	"entangle"	cf.	*za*	

Figure 3.1 Parts of cunieform signs do not reflect their sound or meaning

But also, since Sumerian words were mostly just one syllable long (consonant- or vowel-initial, open or closed), the signs that stood for those syllables could also be used for other similar words for items that could not be easily pictured; one of the earliest examples is the sign for *ti* "arrow" also being used for *ti* "life." (Such reuse is called the *rebus principle*.) As soon as signs came to be used in these transferred ways, they could also be used to record the wide variety of grammatical affixes of Sumerian. The reader could then know the writer's exact intent even when the content was not the stereotyped accounting documents that were, as probably everywhere, the *raison d'être* of the writing system in the first place – even if the writer was not present to explain the text – so that literary and religious compositions of various sorts were soon written down. (The number of such texts never came close to matching in quantity the mundane economic documents.) The vast majority of cuneiform documents record everyday transactions of the widest variety, and clay tablets are close to imperishable (if they have been baked, they are imperishable; if they have only been sun-dried, they can be damaged by water), so that Mesopotamian civilization emerges as the best documented until recent Europe.

The Sumerian language eventually went out of use, to be replaced by the Semitic language Akkadian, but Sumerian remained a language of liturgy and scholarship; and cuneiform writing was used for Akkadian. Akkadian cuneiform is more complicated than Sumerian, because any given sign could have sound value(s) based on its Akkadian meaning(s) as well as its Sumerian, and many syllables could be represented by several different signs, or could be spelled in different ways, and because the Akkadian sound system differs considerably from the Sumerian, and moreover signs could still be used for their meanings rather than their sounds without any indication of such use; in this limited way, a logosyllabic writing system includes isolated instances of purely logographic writing. However, of the 600 or so signs in the Akkadian signlist, only about 200 would be used in any particular time period or area (a selection is shown in table 3.1; the Neo-Assyrian shapes are used in these illustrations). A device for clarifying the writing is the use of *determinatives*, signs (again taken from the normal repertoire) indicating the semantic sphere of the items they accompanied: personal names, wooden objects, cities, countries, plural nouns, etc.

Cuneiform was also used for many other languages of the ancient Near East, such as Elamite, Hurrian, and Urartian, and in these adaptations from Akkadian usage, the script was more syllabic than logographic. An exception is seen in Hittite, which incorporates both Sumerian and Akkadian spellings into texts that nonetheless were to be read in Hittite.

A language that was never written in cuneiform, because it had developed its own writing system, is ancient Egyptian. Rudimentary hieroglyphic writing appears shortly after the beginnings of cuneiform, and it is speculated that the idea of writing somehow came from Sumer to Egypt; but from the very beginning there is no visual similarity and, more important, the sounds recorded are not syllables, but consonants only. Egyptian hieroglyphs remained

Table 3.1 Inventory of basic cuneiform signs used in the pronunciation column of Syllabary A, the signlist studied by Mesopotamian scribal students[a]

	-a	*-e*	*-i*	*-u*	*a-*	*e-*	*i-*	*u-*
p b								
t d ṭ								
k g q								
s z ṣ								
š								
m n								
l r								
w y								
x ʾ								
Ø								

[a] The following CVC signs are also used: *dim* , *dím* , *gír* , *xar* , *kal* , *kil* , *kin* , *kul* , *lag* , *lam* , *rig* , *suk* , *tan* , *tin* (from *WWS*: 57).

recognizable pictures over the 3,500 years they were in use; but from quite early on, a cursive interpretation of them, known as hieratic, was used on papyrus. (*Cursive*: written with speed, character forms affected by the connection of strokes written separately in *formal* or *monumental* styles.) The demotic script emerged considerably later, in connection with a later form of the Egyptian language; there is a one-to-one relationship of hieratic and hieroglyphic signs, but demotic cannot be automatically transposed into the other two scripts.

Egyptian hieroglyphic signs represent one, two, or three consonants (table 3.2). (The monoconsonantal signs were never used as a discrete subsystem for writing Egyptian, so charts of an "Egyptian alphabet" are misleading.) Many signs also function logographically only. Determinatives are used much more

Table 3.2 All the phonetic Egyptian hieroglyphs beginning with labials

b		mḥ		p		wbn	
bꜣ		mỉ		pꜣ		wḏ	
bꜣs		mỉ		pds		wḥm	
bḥ		mn		pḥ		wn	
bỉꜣ		mnw		pr		wn	
bỉꜣ		mr		w		wp	
bỉt		mr		wꜣ		wr	
f		ms		wꜣḏ		wsr	
m		msn		wꜣḥ		wsx	
mꜣ		mt		wꜣs		wsx	
mꜣʿ		mt		wʿ		wšm	
mḏḥ		mw		wʿr			

systematically than in cuneiform, as are *phonetic complements* – signs that give clues to the relevant reading of a logogram. (They are also found in cuneiform.)

Egyptian influence is assumed, but cannot be demonstrated, in the initial development of the writing systems that have spread to all the world except (until recently) East Asia. This development is first certainly seen in the handful of so-called "Proto-Canaanite" inscriptions from the second quarter of the second millennium BCE in the Levant. In the fourteenth century, a 27-letter abjad (with three supplemental letters) clearly standing in the main line of development is well attested at ancient Ugarit (it is written in wedges with a stylus on clay but has no other relation to Mesopotamian cuneiform). This large inventory continued to be used for inscriptions in the South Arabian languages and was taken across to Africa by Sabean colonists who passed it on to the Aksumite kingdom in present-day Ethiopia and Eritrea (by the fourth century CE).

By 1000 BCE or so, a 22-letter script similar to the Hebrew abjad was in wide use (the scattered examples that have been found vary sufficiently to suggest some lengthy period of separate developments; Naveh 1987). Over the next few centuries, indirect methods of indicating vowels developed in Aramaic and Hebrew (but not Phoenician) scribal traditions – to oversimplify, diphthongs (whose glide portions were written with the corresponding consonant letters) contracted into long vowels of related colors, and the consonant letters came to be used for other long vowels as well (*matres lectionis*: "mothers of reading" in Latin), albeit not obligatorily until well into the Common Era, and then only in Mandaic and Arabic. The Aramaic group of scripts tended to cursive developments, one of them surviving in Syriac (Estrangelo and Serto are the principal variants). Another is Nabatean, used by an Arab tribe to write Aramaic and from which a distinctive script for the Arabic language emerged (table 3.3). The Arabic language preserved the full panoply of Proto-Semitic

Table 3.3 West Semitic abjads[a]

Value[b]	*Ugaritic*	*Sabean*[c]	*Phoenician*	*Hebrew*	*Mandaic*	*Estrangelo*	*Serto*	*Arabic*
ʾ (ʾa)	𐎀	𐩱	𐤀	א	ࡀ	ܐ	ܐ	ا a
b	𐎁	𐩨	𐤁	ב	ࡁ	ܒ	ܒ	ب b
g	𐎂	𐩲	𐤂	ג	ࡂ	ܓ	ܓ	ت t
(ḫ)	𐎃	𐩵						ث ṯ
d	𐎄	𐩣	𐤃	ד	ࡃ	ܕ	ܕ	ج j
h {-î}	𐎅	𐩺	𐤄	ה	ࡄ	ܗ	ܗ	ح ḥ
w	𐎆	𐩧	𐤅	ו	ࡅ	ܘ	ܘ	خ x
z	𐎇	𐩬	𐤆	ז	ࡆ	ܙ	ܙ	د d
ḥ {h}	𐎈	𐩦	𐤇	ח	ࡇ	ܚ	ܚ	ذ ḏ
ṭ	𐎉	𐩯	𐤈	ט	ࡈ	ܛ	ܛ	ر r
y	𐎊	𐩠	𐤉	י	ࡉ	ܝ	ܝ	ز z
k	𐎋	𐩫	𐤊	כ	ࡊ	ܟ	ܟ	س s
(š)	𐎌	𐩮						ش š
l	𐎍	𐩡	𐤋	ל	ࡋ	ܠ	ܠ	ص ṣ
m	𐎎	𐩥	𐤌	מ	ࡌ	ܡ	ܡ	ض ḍ
(δ)	𐎏	𐩩						ط ṭ
n	𐎐	𐩤	𐤍	נ	ࡍ	ܢ	ܢ	ظ ẓ
ẓ	𐎑	𐩷						ع ʿ
s	𐎒	𐩪	𐤎	ס	ࡎ	ܣ	ܣ	غ ġ
ʿ	𐎓	𐩢	𐤏	ע	ࡏ	ܥ	ܥ	ف f
p	𐎔	𐩰	𐤐	פ	ࡐ	ܦ	ܦ	ق q
ṣ	𐎕	𐩹	𐤑	צ	ࡑ	ܨ	ܨ	ك k
q	𐎖	𐩭	𐤒	ק	ࡒ	ܩ	ܩ	ل l
r	𐎗	𐩧	𐤓	ר	ࡓ	ܪ	ܪ	م m
ś		𐩸		שׂ				ن n
š (θ)	𐎘	𐩴	𐤔	שׁ	ࡔ	ܫ	ܫ	ه h
(γ)	𐎙	𐩶						و w
t	𐎚	𐩱	𐤕	ת	ࡕ	ܬ	ܬ	ي y
(ʾi)	𐎛							لا lâ
(ʾu)	𐎜							
(š)	𐎝							
(ḍ)		𐩵						
{dî}					ࡖ			

[a] Where two forms are shown, that on the right occurs at the end of a word.
[b] (Ugaritic and Sabean values); {Mandaic values}.
[c] Cf. table 3.11 for the ancient order of the Sabean abjad.

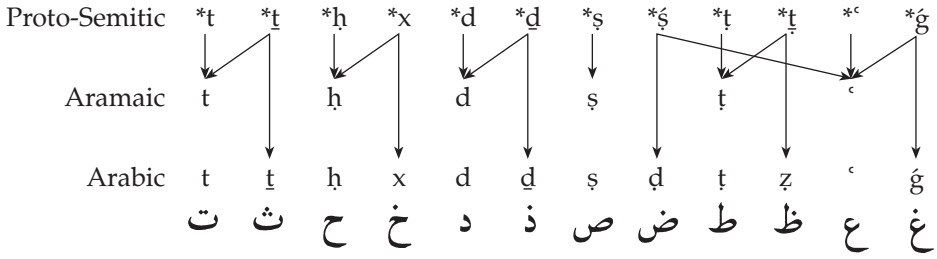

Figure 3.2 Sources of Arabic dotted letters

consonants, and the script includes diacritic dots to distinguish both letters whose shapes merged during its Nabatean prehistory and letters for sounds that had merged in Aramaic but not in Arabic (figure 3.2).

Two script traditions that ultimately left no issue are found at opposite ends of the ancient Near East (table 3.4). Several (logo)syllabic writing systems are found around the Aegean Sea – "Hittite" hieroglyphs (fifteenth to eighth centuries BCE) in western Anatolia, Linear B (sixteenth to thirteenth centuries BCE) in Crete, and Cypriote syllabary (eighth to third centuries BCE) in Cyprus (as well as some presumed antecedents of the latter two, including the still enigmatic Linear A) – for Luvian and two stages of Greek respectively. They are basically pictographic like Egyptian, but they record syllables, not consonants only, and representatives of earlier stages have not been found, so their origin is mysterious. To the east in Iran, a cuneiform script was devised for recording Old Persian (500 BCE). Some of its characters represent syllables, others consonants (probably combining features of the cuneiform and Aramaic scripts that were simultaneously in use in the Persian Empire), but oddly despite its wedge components, it was not used on clay.

A contemporary development in East Asia was the invention of writing for Chinese. While the earliest attested inscriptions (late Shang dynasty, ca. 1200 BCE) are "oracle bone" communications with the gods, most likely writing began there for the same mundane commercial reasons as elsewhere, but only perishable materials were used. The principles of writing Chinese have not changed over more than 3,000 years, though the esthetics and the shapes of the characters certainly have. Earliest written Chinese, like Sumerian, used primarily monosyllabic morphemes, but the combination of phonetic and semantic information was made explicit and obligatory in most "characters" so that the vast majority of characters comprise two parts (table 3.5), and there are considerably more characters in the repertoire. While the biggest dictionaries list upwards of 60,000, an inventory of 5,000 or so characters is adequate for most needs.

Chinese writing was tried for both Korean and Japanese, with unsatisfactory results in both cases. Japanese developed a pair of syllabaries (*kana*) from a selected group of characters that had been in use for their syllabic value.

Table 3.4 The Linear B,[a] Cypriote, and Old Persian[b] syllabaries

	LB	*Cp*	*OP*		*LB*	*Cp*		*LB*	*Cp*	*OP*		*LB*	*Cp*		*LB*	*Cp*	*OP*
a				e			i				o			u			
ba							bi							bu			
ča							či							ču			
ça							çi							çu			
da				de			di				do			du			
fa							fi							fu			
ga										?				gu			
ha							hi							hu			
ja				je			ji				jo			ju			?
ka				ke			ki			?	ko			ku			
la				le			li				lo			lu			
ma				me			mi				mo			mu			
na				ne			ni				no			nu			
pa				pe			pi				po			pu			
qa				qe			qi				qo						
ra				re			ri				ro			ru			
sa				se			si				so			su			
ša							ši							šu			
ta				te			ti				to			tu			
θa							θi							θu			
va							vi										?
wa				we			wi				wo						
xa				xe			xi				xo			xu			
ya							yi							yu			
za				ze			zi				zo			zu			

[a] An additional 16 characters represent variant sounds, and 11 more occur so rarely that they have not been identified. There are also a considerable number of ideograms, identifying commodities, which are not used as logograms in Mycenean Greek prose.

[b] Only the boxed Old Persian characters unambiguously identify the vowel of the syllable; a vowel character is required elsewhere (all *-a* signs used alone stand either for the bare consonant or for the consonant followed by short *a*).

Table 3.5 The construction of Chinese characters (after Gabelentz 1881: 50–1 and DeFrancis 1989: 107; compiled with the assistance of John DeFrancis)

	Phonetic			
Semantic	工 *gōng*	兼 *jiān*	堯 *yāo*	番 *fān*
人 "person"	仜 *hōng* "paunch"	傔 *qiàn* "servant"	僥 *jiǎo* "lucky"	僠 *bō* (a name)
手 "hand"	扛 *káng* "bear"	摲 *jiān* "drum"	撓 *náo* "scratch"	播 *bō* "strew"
水 "water"	江 *jiāng* "stream"	漗 *lián* (a river)	澆 *jiǎo* "sprinkle"	潘 *pān* "ricewater"
系 "silk"	紅 *hóng* "red"	縑 *jiān* "silk cloth"	繞 *rào* "roll up"	繙 *fān* "translate"

Table 3.6 Japanese syllabaries[a]

Hiragana						*Katakana*				
	-a	*-i*	*-u*	*-e*	*-o*	*-a*	*-i*	*-u*	*-e*	*-o*
Ø	あ	い	う	え	お	ア	イ	ウ	エ	オ
k-	か	き	く	け	こ	カ	キ	ク	ケ	コ
g-	が	ぎ	ぐ	げ	ご	ガ	ギ	グ	ゲ	ゴ
s-	さ	し	す	せ	そ	サ	シ	ス	セ	ソ
z-	ざ	じ	ず	ぜ	ぞ	ザ	ジ	ズ	ゼ	ゾ
t-	た	ち	つ	て	と	タ	チ	ツ	テ	ト
d-	だ	ぢ	づ	で	ど	ダ	ヂ	ヅ	デ	ド
n-	な	に	ぬ	ね	の	ナ	ニ	ヌ	ネ	ノ
h-	は	ひ	ふ	へ	ほ	ハ	ヒ	フ	ヘ	ホ
b-	ば	び	ぶ	べ	ぼ	バ	ビ	ブ	ベ	ボ
p-	ぱ	ぴ	ぷ	ぺ	ぽ	パ	ピ	プ	ペ	ポ
m-	ま	み	む	め	も	マ	ミ	ム	メ	モ
y-	や	–	ゆ	–	よ	ヤ	–	ユ	–	ヨ
r-	ら	り	る	れ	ろ	ラ	リ	ル	レ	ロ
w-	わ	–	–	–	を	ワ	–	–	–	ヲ

[a] The syllabic nasal (hi.ん, ka.ン) comes at the end of the list.
Vowel length is indicated in kana by doubling, or more often with a following dash: ああ or あー is *aa*. Geminate consonants are written in kana with a preceding subscript hi.っ, ka.ッ *tu*; thus hi. あっか, ka.アッカ *akka*.

Hiragana are used for writing grammatical morphemes attached to Chinese characters (*kanji*) that are used for content words, and *katakana* are used for foreign words (table 3.6). Korean struggled with characters longer than Japanese and came up with a unique script described below.

Table 3.7 Brahmi-derived scripts of South and Southeast Asia[a]

	Kharoshthi	*Brahmi*	*Devanagari*	*Gujarati*	*Gurmukhi*	*Bengali*	*Oriya*	*Sinhala*	*Kannada*
a	𐨀	𑀅	अ	અ	ਅ	অ	ଅ	අ	ಅ
ā		𑀆	आ	આ	ਆ	আ	ଆ	ආ	ಆ
i	𐨁	𑀇	इ	ઇ	ਇ	ই	ଇ	ඉ	ಇ
ī		𑀈	ई	ઈ	ਈ	ঈ	ଈ	ඊ	ಈ
u	𐨂	𑀉	उ	ઉ	ਉ	উ	ଉ	උ	ಉ
ū		𑀊	ऊ	ઊ	ਊ	ঊ	ଊ	ඌ	ಊ
ṛ		𑀋	ऋ	ઋ		ঋ	ଋ	සෘ	ಋ
ṝ		𑀌	ॠ					සෲ	ೠ
e	𐨏	𑀏	ए	એ	ਏ	এ	ଏ	එ	ಎ
ē								ේ	ಏ
ai		𑀐	ऐ	ઐ	ਐ	ঐ	ଐ	ඓ	ಐ
o	𐨐	𑀑	ओ	ઓ	ਓ	ও	ଓ	ඔ	ಒ
ō								ෝ	ೂ
au			औ	ઔ	ਔ	ঔ	ଔ	ඖ	ಔ
k	𐨑	𑀓	क	ક	ਕ	ক	କ	ක	ಕ
kh	𐨒	𑀔	ख	ખ	ਖ	খ	ଖ	ඛ	ಖ
g	𐨓	𑀕	ग	ગ	ਗ	গ	ଗ	ග	ಗ
gh	𐨔	𑀖	घ	ઘ	ਘ	ঘ	ଘ	ඝ	ಘ
ṅ		𑀗	ङ		ਙ	ঙ	ଙ	ඞ	ಙ
c	𐨕	𑀘	च	ચ	ਚ	চ	ଚ	ච	ಚ
ch	𐨖	𑀙	छ	છ	ਛ	ছ	ଛ	ඡ	ಛ
j	𐨗	𑀚	ज	જ	ਜ	জ	ଜ	ජ	ಜ
jh	𐨘	𑀛	झ	ઝ	ਝ	ঝ	ଝ	ඣ	ಝ
ñ	𐨙	𑀜	ञ		ਞ	ঞ	ଞ	ඤ	ಞ
ṭ	𐨚	𑀝	ट	ટ	ਟ	ট	ଟ	ට	ಟ
ṭh	𐨛	𑀞	ठ	ઠ	ਠ	ঠ	ଠ	ඨ	ಠ
ḍ	𐨜	𑀟	ड	ડ	ਡ	ড	ଡ	ඩ	ಡ
ḍh	𐨝	𑀠	ढ	ઢ	ਢ	ঢ	ଢ	ඪ	ಢ
ṇ	𐨞	𑀡	ण	ણ	ਣ	ণ	ଣ	ණ	ಣ
t	𐨟	𑀢	त	ત	ਤ	ত	ତ	ත	ತ
th	𐨠	𑀣	थ	થ	ਥ	থ	ଥ	ථ	ಥ
d	𐨡	𑀤	द	દ	ਦ	দ	ଦ	ද	ದ
dh	𐨢	𑀥	ध	ધ	ਧ	ধ	ଧ	ධ	ಧ
n	𐨣	𑀦	न	ન	ਨ	ন	ନ	න	ನ
p	𐨤	𑀧	प	પ	ਪ	প	ପ	ප	ಪ
ph	𐨥	𑀨	फ	ફ	ਫ	ফ	ଫ	ඵ	ಫ
b	𐨦	𑀩	ब	બ	ਬ	ব	ବ	බ	ಬ
bh	𐨧	𑀪	भ	ભ	ਭ	ভ	ଭ	භ	ಭ
m	𐨨	𑀫	म	મ	ਮ	ম	ମ	ම	ಮ
y	𐨩	𑀬	य	ય	ਯ	য়	ଯ	ය	ಯ
r	𐨪	𑀭	र	ર	ਰ	র	ର	ර	ರ
l	𐨫	𑀮	ल	લ	ਲ	ল	ଲ	ල	ಲ
v	𐨬	𑀯	व	વ	ਵ		ୱ	ව	ವ
ś	𐨭	𑀰	श	શ		শ	ଶ	ශ	ಶ
ṣ	𐨮	𑀱	ष	ષ		ষ	ଷ	ෂ	ಷ
s	𐨯	𑀲	स	સ	ਸ	স	ସ	ස	ಸ
h	𐨰	𑀳	ह	હ	ਹ	হ	ହ	හ	ಹ
ḷ			ळ	ળ				ළ	ೞ

Table 3.7 cont'd

	Telugu	*Malayalam*	*Tamil*	*Tibetan*	*Burmese*	*Thai*	*Lao*	*Khmer*	*Javanese*
a	అ	അ	அ	ཨ	အ	อ	ອ	អ	ꦄ
ā	ఆ	ആ	ஆ	ཨཱ	အာ	อา		អា	
i	ఇ	ഇ	இ					ឥ	ꦆ
ī	ఈ	ഈ	ஈ					ឦ	
u	ఉ	ഉ	உ					ឧ	ꦈ
ū	ఊ	ഊ	ஊ					ឩ	
ṛ	ఋ	ഋ							
ṝ	ౠ								
e	ఎ	എ	எ						ꦌ
ē	ఏ	ഏ	ஏ					ឯ	
ai	ఐ	ഐ	ஐ					ឰ	
o	ఒ	ഒ	ஒ					ឱ	ꦎ
ō	ఓ	ഓ	ஓ						
au	ఔ	ഔ	ஔ						
k	క	ക	க	ཀ	က	ก	ກ	ក	ꦏ
kh	ఖ	ഖ		ཁ	ခ	ข ฃ	ຂ	ខ	
g	గ	ഗ		ག	ဂ	ค ฅ	ຄ	គ	ꦒ
gh	ఘ	ഘ			ဃ	ฆ		ឃ	
ṅ		ങ	ங	ང	င	ง	ງ	ង	ꦔ
c	చ	ച	ச	ཅ	စ	จ	ຈ	ច	ꦕ
ch	ఛ	ഛ		ཆ	ဆ	ฉ	(ສ)	ឆ	
j	జ	ജ	ஜ	ཇ	ဇ	ช ฌ	ຊ	ជ	ꦗ
jh	ఝ	ഝ			ဈ	ฌ		ឈ	
ñ	ఞ	ഞ	ஞ	ཉ	ည	ญ	ຍ	ញ	ꦚ
ṭ	ట	ട	ட	ཊ	ဋ	ฏ ฎ		ដ	ꦝ
ṭh	ఠ	ഠ		ཋ	ဌ	ฐ		ឋ	
ḍ	డ	ഡ		ཌ	ဍ	ฑ		ឌ	ꦣ
ḍh	ఢ	ഢ		ཌྷ	ဎ	ฒ		ឍ	
ṇ	ణ	ണ	ண	ཎ	ဏ	ณ		ណ	
t	త	ത	த	ཏ	တ	ต ฏ	ຕ ຖ	ត	ꦠ
th	థ	ഥ		ཐ	ထ	ถ	ຖ	ថ	
d	ద	ദ		ད	ဒ	ท	ທ	ទ	ꦢ
dh	ధ	ധ			ဓ	ธ		ធ	
n	న	ന	ந	ན	န	น	ນ	ន	ꦤ
p	ప	പ	ப	པ	ပ	บ ป	ບ ປ	ប	ꦥ
ph	ఫ	ഫ		ཕ	ဖ	ผ ฝ	ຜ ຝ	ផ	
b	బ	ബ		བ	ဗ	พ ฟ	ພ ຟ	ព	ꦧ
bh	భ	ഭ			ဘ	ภ		ភ	
m	మ	മ	ம	མ	မ	ม	ມ	ម	ꦩ
y	య	യ	ய	ཡ	ယ	ย	ຢ	យ	ꦪ
r	ర	ര	ர	ར	ရ	ร	ຣ ຮ	រ	ꦫ
l	ల	ല	ல	ལ	လ	ล	ລ	ល	ꦭ
v	వ	വ	வ	ཝ	ဝ	ว	ວ	វ	ꦮ
ś	శ	ശ		ཤ		ศ			
ṣ	ష	ഷ	ஷ	ཥ		ษ			
s	స	സ	ஸ	ས	သ	ส	ສ	ស	ꦯ
h	హ	ഹ	ஹ	ཧ	ဟ	ห	ຫ	ហ	ꦲ
ḷ	ళ	ള	ள		ဠ	ฬ		ឡ	ꦭ꦳

[a] Some letters in only one or a few scripts, e.g. the Sinhala prenasalized series, are omitted. Phonetic values of letters may not be exact, especially in later (rightward in table) scripts.

1.2 From abjad to alphabet and abugida

Abjads seem well suited to Semitic languages, which are supposed to involve consonantal "roots" and vowel "patterns" (though this analysis is increasingly recognized as an artefact of the Arabic writing system as it was available to the Arab grammarians who devised it), but are less appropriate to Indo-European languages where vocalization is more unpredictable than in Semitic. Two different schemes for the obligatory recording of vowels emerged. The first, seen with the first attempts to write Greek with the Phoenician abjad, probably around 800 BCE, seems accidental and inevitable: Semitic has a larger repertoire of consonants than Greek, and (phonemic perception being what it is), the letters representing sounds, especially laryngeals, not found in Greek would be heard as indicating the succeeding vowels. Thus Phoenician ‹ʾ› was taken to represent /a/, ‹h› for /e/, ‹y› for /i/, ‹ʿ› for /o/, and ‹w› for /u/. And so the alphabet was born. The correspondences are seen in both the shapes of the letters and their positions in the respective alphabetical orders. Greek settlements used slightly varying inventories of letters; the most significant for the history of writing was in Italy, where the alphabet was passed on to the Etruscans and other local peoples, and in turn from the Etruscans to the Romans.

The second Indo-European adaptation of the Semitic abjad occurred in India (table 3.7), probably no earlier than the third century BCE (Falk 1993). Here the method was not separate letters for vowels, but appendages – left, right, above, or below – to the consonant letters to designate the vowels (short other than *a*, and long) and diphthongs of the Indic and Dravidian languages, using the type I call abugida (table 3.8). The first language written with the Kharoṣṭhi and Brahmi scripts was Prakrit (a colloquial variety that developed later than the Sanskrit "literary" language); it was several centuries before it became licit to write down the sacred Sanskrit texts that had been preserved orally for hundreds of years. Limited communication between the sundry regions and polities of India, as well as differences in writing materials, led to considerable diversity in appearance of the script, and Sanskrit texts as well as local languages would be written in each locality's distinctive hand, but the abugidic principle remained uniform. A consonant-final word had a mark to indicate that the final letter was pronounced vowelless, and immediately adjacent consonants (whether in the same syllable or not) were written by combining reduced forms of the consonant letters into a single symbol (figure 3.3). Today, ten standardized scripts serve the Indian subcontinent's literary Indo-Aryan and Dravidian languages.

The Indic style of writing was carried by Buddhist missionaries throughout Southeast Asia, where essentially the same principle remains at work in such diverse-looking scripts as the Thai and Lao, Burmese, Khmer (Cambodian), and Javanese, as well as a host of less standardized ones. The missionaries also brought writing to Tibet late in the first millennium CE (apparently from southern India, though the lineage of the Tibetan script is not entirely clear).

Table 3.8 Vowel indications in some scripts of South and Southeast Asia

		a	*ā*	*i*	*ī*	*u*	*ū*	*e*	*ai*	*o*	*au*
Brahmi	k	𑀓	𑀓𑀸	𑀓𑀺	𑀓𑀻	𑀓𑀼	𑀓𑀽	𑀓𑁂	𑀓𑁃	𑀓𑁄	𑀓𑁅
	g	𑀕	𑀕𑀸	𑀕𑀺	𑀕𑀻	𑀕𑀼	𑀕𑀽	𑀕𑁂	𑀕𑁃	𑀕𑁄	𑀕𑁅
Devanagari	k	क	का	कि	की	कु	कू	के	कै	को	कौ
	g	ग	गा	गि	गी	गु	गू	गे	गै	गो	गौ
Oriya	k	କ	କା	କି	କୀ	କୁ	କୂ	କେ	କୈ	କୋ	କୌ
	g	ଗ	ଗା	ଗି	ଗୀ	ଗୁ	ଗୂ	ଗେ	ଗୈ	ଗୋ	ଗୌ
Javanese	k	ꦏ		ꦏꦶ		ꦏꦸ		ꦏꦺ		ꦏꦺꦴ	
	g	ꦒ		ꦒꦶ		ꦒꦸ		ꦒꦺ		ꦒꦺꦴ	

Here, though, consonant clusters were not notated by combining symbols into a single character regardless of syllable division. Rather, the end of every syllable is marked with a dot, so that syllable-final consonants are kept separate from syllable-initial consonants, while vowels are still indicated by additions for *e*, *i*, *o* (above), and *u* (below).

Missionaries had been active in Europe as well. The Roman alphabet accompanied the Roman church with its Latin liturgy throughout western Europe, but in the Eastern church, where the liturgy was conducted in local languages, separate scripts were devised for a number of languages (table 3.9) – among them Coptic, Gothic, Armenian, and Georgian (fourth century), and, for Old Church Slavonic, Glagolitic and Cyrillic (ninth century). Coptic and Gothic are adaptations of the Greek alphabet; as is seen by the inherited letter order, the next two are inspired by it, though the shapes of the letters seem to be independent creations; the last two appear to be based on cursive and formal Greek writing respectively. All these alphabets except Gothic require considerably more letters than the Greek (Coptic's additions are taken from demotic Egyptian script). Northern and northwestern Europe saw local developments of runes

Brahmi	Devanagari	
𑀔 *kha* + 𑀬 *ya* = 𑀔𑁆𑀬 *khya*	क *ka* + ल *la* = क्ल *kla*	क *ka* + ष *ṣa* = क्ष *kṣa*
𑀧 *pa* + 𑀢 *ta* = 𑀧𑁆𑀢 *pta*	ह *ha* + न *na* = ह्न *hna*	क्ष *kṣa* + म *ma* = क्ष्म *kṣma*
Oriya	**Javanese**	
ଘ *gha* + ନ *na* = ଘ୍ନ *ghna*	ꦤ *na* + ꦭ *la* = ꦤ꧀ꦭ *nla*	ꦭ *la* + ꦲ *ha* = ꦭ꧀ꦲ *lha*
ସ *sa* + ଥ *tha* = ସ୍ଥ *stha*	ꦚ *ña* + ꦧ *ba* = ꦚ꧀ꦧ *ñba*	ꦢ *da* + ꦤ *na* = ꦢ꧀ꦤ *dna*
ଦ *da* + ଧ *dha* = ଦ୍ଧ *ddha*	ꦧ *ba* + ꦚ *ña* = ꦧ꧀ꦚ *bña*	

Figure 3.3 Some consonant clusters in South and Southeast Asian scripts

Table 3.9 Alphabets derived from Greek

Greek		*Coptic*	*Gothic*	*Armenian*		*Georgian*[a]		*Glagolitic*	*Old Cyrillic*	*Russian*	
Α	α	ⲁ	𐌰	a	ա	a	ა	a	а	А	а
Β	β	ⲃ	𐌱	b	բ	b	ბ	b	б	Б	б
Γ	γ	ⲅ	𐌲	g	գ	g	გ	v	в	В	в
Δ	δ	ⲇ	𐌳	d	դ	d	დ	g	г	Г	г
Ε	ε	ⲉ	𐌴	e	ե	e	ე	d	д	Д	д
Ϝ	(6)		𐌵	z	զ	v	ვ	e	є	Е	е
Ζ	ζ	ⲍ	𐌶	ē	է	z	ზ	ž	ж	Ж	ж
Η	η	ⲏ	𐌷	ə	ը	ey	‹ეი›	dz	ѕ		
Θ	θ	ⲑ	𐌸	t′	թ	t	თ	z	з	З	з
Ι	ι	ⲓ	𐌹	ž	ժ	i	ი	i	и	И	и
Κ	κ	ⲕ	𐌺	i	ի	ķ	კ	i	ї/і	Й	й
Λ	λ	ⲗ	𐌻	l	լ	l	ლ	ģ	ѓ		
Μ	μ	ⲙ	𐌼	x	խ	m	მ	k	к	К	к
Ν	ν	ⲛ	𐌽	c	ծ	n	ნ	l	л	Л	л
Ξ	ξ	ⲝ	𐌾	k	կ	y	‹ჲ›	m	м	М	м
Ο	ο	ⲟ	𐌿	h	հ	o	ო	n	н	Н	н
Π	π	ⲡ	𐍀	j	ձ	ṗ	პ	o	о	О	о
Ϙ	(90)		𐍁	ł	ղ	ž	ჟ	p	п	П	п
Ρ	ρ	ⲣ	𐍂	č	ճ	r	რ	r	р	Р	р
Σ	σ ς	ⲥ	𐍃	m	մ	s	ს	s	с	С	с
Τ	τ	ⲧ	𐍄	y	յ	ṭ	ტ	t	т	Т	т
Υ	υ	ⲩ	𐍅	n	ն	wi	‹ჳ›	u	оу/ѹ	У	у
Φ	ϕ	ⲫ	𐍆	š	շ	u	უ	f	ф	Ф	ф
Χ	χ	ⲭ	𐍇	o	ո	p	ფ	x	х	Х	х
Ψ	ψ	ⲯ	𐍈	č′	չ	k	ქ	o	ѡ/ѿ		
Ω	ω	ⲱ	𐍉	p	պ	ğ	ღ	c	ц	Ц	ц
Ϡ	(900)		𐍊	ǰ	ջ	q	ყ	č	ч	Ч	ч
		š ⲋ		ŕ	ռ	š	შ	š	ш	Ш	ш
		f ⲫ		s	ս	č	ჩ	št	щ	Щ	щ
		h ⲭ		v	վ	c	ც	″/ъ	ъ	Ъ	ъ
		j ϫ		t	տ	j	ძ	y	ы/ъи	Ы	ы
		q ϭ		r	ր	ç	წ	′/ь	ь	Ь	ь
		ti ϯ		c′	ց	č	ჭ	ě	ѣ	Э	э
				w	ւ	x	ხ	ju	ю	Ю	ю
				p′	փ	q	‹ყ›	ja	ꙗ	Я	я
				k′	ք	ǰ	ჯ	je	є		
				u	ու	h	ჰ	ę	ѧ		
				ō	օ	ow	‹ფ›	ję	ѩ		
				f	ֆ			ǫ	ѫ		
								jǫ	ѭ		
								ks	ѯ		
								ps	ѱ		
								f	ѳ		
								i/v	ѵ		

[a] ‹Obsolete letters, no longer used in Georgian›.

Table 3.10 Runes, Ogham

	Runes				*Ogham*		
f	ᚠ	*fehu	“wealth”	b	ᚁ	Beithe	“birch”
u	ᚢ	*ūruz	“aurochs”	l	ᚂ	Luis	“blaze”
þ	ᚦ	*þurisaz	“giant”	f	ᚃ	Fern	“alder”
a	ᚨ	*ansuz	“god”	s	ᚄ	Sail	“willow”
r	ᚱ	*raiþō	“riding”	n	ᚅ	Nin	“fork”
k	ᚲ	*kaunaz	“ulcer”	h	ᚆ	(h)Úath	“fear”(?)
		*kēnaz	“torch”	d	ᚇ	Dair	“oak”
		*kanō	“skiff”	t	ᚈ	Tinne	“metal rod”
g	ᚷ	*geƀō	“gift”	c	ᚉ	Coll	“hazel”
w	ᚹ	*wunjō	“joy”	q	ᚊ	Queirt	“bush”
h	ᚺ	*hagalaz	“hail”	m	ᚋ	Muin	“neck”
n	ᚾ	*nauþiz	“need”	g	ᚌ	Gort	“field”
i	ᛁ	*isa-	“ice”	ng	ᚍ	(n)Gétal	“wounding”(?)
j	ᛃ	*jēra-	“year”	z	ᚎ	Straif	“sulfur”
ï	ᛇ	*eihwaz	“yew tree”	r	ᚏ	Ruis	“red(ness)”
p	ᛈ	*perþ-	?	a	ᚐ	Ailm	?
z	ᛉ	*algiz	“sedge” (?)	o	ᚑ	Onn	“ash-tree”
s	ᛋ	*sōwulō	“sun”	u	ᚒ	Úr	“earth”
t	ᛏ	*teiwaz	god Tyr	e	ᚓ	Edad	?
b	ᛒ	*ƀerkana-	“birch twig”	i	ᚔ	Idad	?
e	ᛖ	*ehwaz	“horse”	ea	ᚕ	Ébad	?
m	ᛗ	*mannaz	“man”	oi	ᚖ	Ór	“gold”
l	ᛚ	*laguz	“water”	ui	ᚗ	Uilen	“elbow”
ng	ᛜ	*inguz	god Ing	ia	ᚘ	Pín	“pine”
d	ᛞ	*đagaz	“day”	ae	ᚙ	Emancholl	“double c”
o	ᛟ	*ōþila	“inherited land”				

for Germanic languages (first to ninth centuries) and Ogham for Irish (fifth to seventh centuries); both reveal the influence of the Latin alphabet (table 3.10).

Meanwhile, the Semitic scripts could not go forever with no means of explicitly indicating vowels. The first one to innovate such a device was the Ethiopic (table 3.11), which became an abugida suddenly at the same time as the country adopted Christianity (ca. 350); apparently some knowledge of Indic writing was involved, though the shapes of the vowel indicators are not similar in the two systems, and the basic consonants do not retain their shapes so rigorously as in India. Perhaps the ancient Christian communities of western India supplied personnel for the trading voyages that regularly crossed the Arabian Sea with the monsoons, and the idea, though not the details, of Indic writing went with them.

Table 3.11 The Ethiopic abugida (ancient Sabean order for comparison)

Value	*-ä*	*-u*	*-i*	*-a*	*-e*	*-ə*	*-o*	*Sabean*	
h	ሀ	ሁ	ሂ	ሃ	ሄ	ህ	ሆ	h	𐩱
l	ለ	ሉ	ሊ	ላ	ሌ	ል	ሎ	l	𐩣
ḥ	ሐ	ሑ	ሒ	ሓ	ሔ	ሕ	ሖ	ḥ	𐩧
m	መ	ሙ	ሚ	ማ	ሜ	ም	ሞ	m	𐩬
š	ሠ	ሡ	ሢ	ሣ	ሤ	ሥ	ሦ	q	𐩮
r	ረ	ሩ	ሪ	ራ	ሬ	ር	ሮ	w	𐩵
s	ሰ	ሱ	ሲ	ሳ	ሴ	ስ	ሶ	s^2	𐩺
q	ቀ	ቁ	ቂ	ቃ	ቄ	ቅ	ቆ	r	𐩲
q^w	ቈ		ቊ	ቋ	ቌ	ቍ		b	𐩫
b	በ	ቡ	ቢ	ባ	ቤ	ብ	ቦ	t	𐩸
t	ተ	ቱ	ቲ	ታ	ቴ	ት	ቶ	s^1	𐩦
x	ኀ	ኁ	ኂ	ኃ	ኄ	ኅ	ኆ	k	𐩩
x^w	ኈ		ኊ	ኋ	ኌ	ኍ		n	𐩭
n	ነ	ኑ	ኒ	ና	ኔ	ን	ኖ	x	𐩤
ʾ	አ	ኡ	ኢ	ኣ	ኤ	እ	ኦ	ṣ	𐩥
k	ከ	ኩ	ኪ	ካ	ኬ	ክ	ኮ	s^3	𐩯
k^w	ኰ		ኲ	ኳ	ኴ	ኵ		f	𐩠
w	ወ	ዉ	ዊ	ዋ	ዌ	ው	ዎ	ʾ	𐩡
ʿ	ዐ	ዑ	ዒ	ዓ	ዔ	ዕ	ዖ	ʿ	𐩢
z	ዘ	ዙ	ዚ	ዛ	ዜ	ዝ	ዞ	ḍ	𐩰
y	የ	ዩ	ዪ	ያ	ዬ	ይ	ዮ	g	𐩪
d	ደ	ዱ	ዲ	ዳ	ዴ	ድ	ዶ	d	𐩨
g	ገ	ጉ	ጊ	ጋ	ጌ	ግ	ጎ	γ	𐩳
g^w	ጐ		ጒ	ጓ	ጔ	ጕ		ṭ	𐩴
ṭ	ጠ	ጡ	ጢ	ጣ	ጤ	ጥ	ጦ	z	𐩶
ṗ	ጰ	ጱ	ጲ	ጳ	ጴ	ጵ	ጶ	δ	𐩷
ṣ	ጸ	ጹ	ጺ	ጻ	ጼ	ጽ	ጾ	y	𐩹
ḍ	ፀ	ፁ	ፂ	ፃ	ፄ	ፅ	ፆ	θ	𐩻
f	ፈ	ፉ	ፊ	ፋ	ፌ	ፍ	ፎ	ẓ	𐩼
p	ፐ	ፑ	ፒ	ፓ	ፔ	ፕ	ፖ		

Syriac was the first Semitic script to add optional symbols with the effect of denoting vowels: at first they were a single dot that marked a "fuller" vowel above the consonant it followed and a weaker vowel below; these developed into, on the one hand, markers of grammatical categories, and on the other, markers of vowel quality. This system has survived to the present in Eastern Syriac. In the western area, the optional symbols were Greek vowel letters, written small, that could be placed above or below their consonants. In Hebrew, several scholarly circles devised different sets of marks for indicating vowel quality, prosodic and syntactic characteristics of the text, and liturgical

Table 3.12 West Semitic vowel signs (shown with the consonant *b*)

	Hebrew	*Syriac (Eastern)*	*Syriac (Western)*	*Arabic*
i	בִ	ܒ݂	ܒ݁, ܒ݂	بِ (ـٍ *-in*)
e	בֵ	ܒ݁		
ɛ	בֶ	ܒ݂	ܒ݂, ܒ݁	
a	בַ	ܒ݁	ܒ݁, ܒ݁	بَ (ـً *-an*)
ɔ	בָ	ـ ـ (*ā*)		ا (*ʾā*)
o	בֹ	ـ ـ	ـ ـ, ـ ـ	
u	בֻ, בוּ–	ـ ـ	ـ ـ, ـ ـ	بُ (ـٌ *-un*)
Ø	בְ			بْ
ə	בְ			
ĕ	בֱ			
ă	בֲ			
ŏ	בֳ			

melodies (the Tiberian system is the only one still in use). Arabic marks the three short vowels and a number of morphophonemic phenomena (with, as mentioned, all long vowels obligatorily notated within the line of consonantal letters). In all three languages (table 3.12), a major impetus for adding vowel (and accentual and musical) notation ("pointing") outside the consonantal text was the preexisting text of Scripture, which needed to be preserved in full detail; an explanation for the complications of Arabic is that the pointing seems to have been added by speakers of a different dialect from that used by the scribes who recorded the consonantal text of the Qur'ān (Versteegh 1997: 56).

While vowel notations were being devised for the major Semitic literary languages, the same was not happening for the Iranian languages that gradually adopted Aramaic writing, specifically the Manichean script: among others, in the west Parthian and Pahlavi, in the east Sogdian. Rather than a brusque adaptation, as described for Greek and Prakrit, the Iranian scribes apparently continued to keep their records in Aramaic, but as knowledge of Aramaic deteriorated, Iranian forms crept in. Eventually a system developed whereby many words that were pronounced as Iranian continued to be written with the Aramaic spellings – but with grammatical affixes spelled in proper Iranian. In effect, Pahlavi and so on were written logoconsonantally (Skjærvø 1997). (Moreover, a number of letters merged in shape, making these texts very difficult to read letter by letter, so the logogram gestalts are better than Iranian spellings would be.)

When it came time, however, to preserve in writing the oral tradition of the Zoroastrian scripture, the Avesta (which was in danger of being lost because the language was no longer clearly understood – the texts were preserved

Table 3.13 Manichean, Parthian, Pahlavi, Avestan, Sogdian[a]

Manichean < Aramaic	*Parthian Inscrips*	*Pahlavi Psalter*	*Book Pahlavi*	*Avestan*	*Source of Avestan*[b]	*Sogdian*
ʾ	a, ā	a, ā	a, ā	a	Phl.	a, ā
				h	Av. *a*	
				x	Av. *h*	
				ā	Phl. *ay*	
				å, ā̊	Av. *āə*	
				ą, ą̇	?	
b	b, w		b, w	b	Phl.	b, β
g	g, γ		g, y	g, ġ	Phl.?	g, γ
d	d, δ		d, y	d	Ps.	d, δ
				δ	?	
h (ḫ)				ə, ə̄	Gk. ε	a, Ø
w	w, ŏ, ŭ	ʾ = ʿ, r	w, ŏ, ŭ	u, ū	Phl.	w, ŏ, ŭ
z	z, ž		z	z	Phl.	z
ḥ (h)	h, x		h, x	x^{v}	Phl. *ḥw*	γ, x, h
ṭ				ẋ	Av. *x^{v}*	
y	y, ĕ, ĭ		y, ĕ, ĭ, ǰ	i, ī	Phl.	y, ĕ, ĭ
				e, ē	Phl. *ēw*	
k	k, g		k, g	k	Phl.	k
				γ	Ps. k	
l (δ)	l		l, r	r	Phl.	δ
				o	Phl. *(ʿ)l* [ō]	
				ō	Phl. *ʿl* [ō]	
m	m		m	m, ṃ	Phl.	m
n	n		n	n	Phl.	n
				ń	Av. *n*	
				ŋ, ŋ́, ŋ́^{v}, ṇ	?	
s	s		s, h	s	Phl.	s
				θ	Phl. *s*	
ʿ	= r	ʾ = w, r	Ø			Ø
p	p, b		p, b, f	p, f	Phl.	p
				β	Av. *p*	
ṣ (c)	č		č, ǰ, z	j	Ps.	č, ǰ
				c	Av. *j*	
q				ž	Phl. *ʾc*	
r	r = ʿ	ʾ = w, ʿ	r			r
š	š, ž		š	š, ś, y, ṣ	Phl.	š
t	t, d		t, d	t	Phl.	t, θ
				ṭ	Av. *t*	
				ẏ, v	?	

[a] Iranian fonts courtesy of P. Oktor Skjærvø, Harvard University.
[b] According to Hoffmann 1988, summarizing Hoffmann and Narten 1989.

Table 3.14 Uyghur, Mongolian, Manchu (after Coulmas 1996: 526, 344, 322)

Uyghur		*Mongolian*		*Manchu*			
ʾ		a		a		s	
β		e		e		dz	
γ		i		i		ts	
w		o u u̱		o		š	
z		ö ü ü̱		u		ž	
x		n		ū		t̲	
y		ng (ŋ)		ã		d̲	
k		q		n		t	
d(δ)		ġ γ		k		d	
m		b		g		tš	
n		p		χ		tšh	
s		s		ḱ		dž	
p		š		ǵ		džh	
č		t d		χ́		l	
r		l		ḱh		r	
š		m		ǵh		m	
t		č		χ́h		y	
l		j		b		v	
ž		y i̱		p		f	
-m		k g ǵ					
q̋		r					
		v					
		h					

purely as long stretches of sound), a new Avestan alphabet was devised (fifth to sixth centuries CE) that used the shapes of the earlier Aramaic-based Iranian consonant letters, added new ones built on them, and thanks to familiarity with Greek, included letters for many vowels (table 3.13).

With the spread of Inner Asian polities, Altaic languages came into contact with Iranian ones: Turkic Uyghur, Mongolic Mongolian, and Tungusic Manchu in turn adopted (but scarcely adapted) Iranian writing, specifically Sogdian (turning it vertical to conform to the Chinese esthetic). This old Mongolian script is being revived in newly democratic Mongolia (table 3.14).

The Mongolian emperor Kubla Khan (thirteenth century), recognizing the inadequacy of Mongolian script for the variety of languages used within his realm, commissioned the Tibetan monk hPags pa to create a script to be used for Mongolian, Tibetan, Chinese, etc. (in the event, it was used primarily for Mongolian), and the script that bears hPags pa's name is modeled closely on Tibetan as to shape (though severely squared up) but is written in columns; it retains the abugidic principle, but places all the non-*a* vowel indicators after (i.e. below) the consonant they follow, and gives up indication of syllable boundaries (table 3.15).

Acquaintance with the hPags-pa script and deep familiarity with the Chinese grammatical tradition, as has been demonstrated by careful philological investigation (Kim-Renaud 1997), underlie the Korean alphabet, promulgated by King Sejong in 1443. It goes beyond both of these, however, in recognizing the separate existence of syllable-final consonants (as Chinese theory did not), identifying them with the initials that had been recognized by Chinese grammarians. Consonants and vowels receive very differently shaped symbols: the basic consonant signs are explicitly iconic representations of the vocal tract involved in producing each, and the basic vowel signs relate to the fundamental principles "heaven," "earth," and "man." Korean is thus a featural script (Kim 1980 [1988], Sampson 1985); and the consonants and vowel of each syllable are written within a square space, in imitation of Chinese characters, so that it is featural, alphabetic, and syllabic all at once (figure 3.4). Three of the scripts of western Eurasia have been adapted to write many languages during the last millennium or so: Roman-based alphabets and Arabic-based alphabets and abjads tend to account for new sounds by adding diacritics to existing letters, while Cyrillic-based ones tend to add new lettershapes.

1.3 Logosyllabaries of the New World and syllabaries of the modern world

Outside the two great old world families of writing systems, the Semitic-derived and the Chinese-derived, which converge in Korean writing, two further phenomena must be mentioned. First, in Meso-America, a large number of inscriptions are known, in upwards of a dozen different forms of writing or proto-writing. The interrelations of these systems are still being puzzled out, but the best understood one, the Maya hieroglyphs (perfected by the ninth century CE), has proven to be a logosyllabic script quite similar in structure to Sumerian (table 3.16).

Table 3.15 Tibetan, hPags pa, Korean

	Ti	*hPp*	*Ko*		*Ti*	*hPp*	*Ko*		*Ti*	*hPp*	*Ko*		*Ti*	*hPp*	*Ko*
k	ཀ		ㄱㄱ	kh	ཁ		ㅋ	g	ག		ㄱ	n̄	ང		ㅇ
				q				γ							
c	ཅ		ㅉㅉ	ch	ཆ		ㅊ	j	ཇ		ㅈ	ñ	ཉ		
t	ཏ		ㄸㄸ	th	ཐ		ㅌ	d	ད		ㄷ	n	ན		ㄴ
p	པ		ㅃㅃ	ph	ཕ		ㅍ	b	བ		ㅂ	m	མ		ㅁ
ts	ཙ			tsh	ཚ			dz	ཛ						
(ʼa)	ཝ			ž	ཞ			z	ཟ						
w	ཝ			y	ཡ			r	ར			l	ལ		ㄹ
u̯				i̯											
f	ཕ			š	ཤ			s	ས		ㅅ	ss			ㅆ
h	ཧ		ㅎ	(a)	ཨ										
i	ི		ㅣ	u	ུ		ㅡ	e	ེ		ㅓ	o	ོ		ㅗ
ey			ㅖ									wu			ㅜ
ay			ㅐ									a			ㅏ

Second, there are upwards of a dozen cases of scripts independently devised in modern times – invented by people who could not read or write in any language, but simply were aware of the existence of writing (usually that of Christian or Muslim missionaries). Earliest and most familiar is the Cherokee syllabary, devised in the 1820s by Sequoyah (table 3.17A). Over the next century or so, a number of syllabaries were invented in Africa, as well as some in North and South America and in Oceania.

This rapid survey of the world's writing systems closes with mention of scientifically created scripts, informed by phonetic science. Noteworthy is the Cree script of the Methodist missionary James Evans of the early 1840s (it and adaptations are used for several languages of Canada); it is featural-cum-abugida (table 3.17B). The two prominent shorthand systems, Pitman (1837) and Gregg (1888), are featural. So are some scripts devised by phoneticians for close recording of speech, but none of them remained in use; the International Phonetic Alphabet and similar systems used by various language specialists are in effect greatly extended alphabets, with featural diacritics available for additional subtlety (Pullum and Ladusaw 1996).

바 *pa* "rope"	밤 *pam* "night"
소 *so* "cow"	손 *so* "hand"
읽다 *ilkta* "read"	밟다 *palpta* "tread on"

Figure 3.4 Korean syllable formation

Table 3.16 The Maya syllabary (Coulmas 1996: 332–3)

	a	**e**	**i**	**o**	**u**
					
b					
ch					
ch′					
h					
c					
k					
l					
m					
n					
p					
s					
t					
tz					
dz					
u					
x					
y					

Table 3.17 The Cherokee and Cree syllabaries

A. Cherokee

-a						*-e*				*-i*				*-o*		*-u*		*-v* = [ə]	
a	Ꭰ					e	Ꭱ			i	Ꭲ			o	Ꭳ	u	Ꭴ	v	ᵢ
ga	Ꭶ	ka	Ꭷ			ge	Ꭸ			gi	Ꭹ			go	Ꭺ	gu	Ꭻ	gv	ᵿ
ha	Ꭼ					he	Ꭽ			hi	Ꭾ			ho	Ꭿ	hu	Ꮀ	hv	ᵛ
la	Ꮁ					le	Ꮂ			li	Ꮃ			lo	Ꮄ	lu	Ꮅ	lv	Ꮹ
ma	Ꮆ					me	Ꮇ			mi	Ꮈ			mo	Ꮉ	mu	Ꮊ		
na	Ꮋ	hna	Ꮌ	nah	Ꮍ	ne	Ꮎ			ni	Ꮏ			no	Ꮐ	nu	Ꮑ	nv	Ꮻ
qua	Ꮒ					que	Ꮓ			qui	Ꮔ			quo	Ꮕ	quu	Ꮖ	quv	Ᶎ
s	Ꮗ	sa	Ꮘ			se	Ꮙ			si	Ꮚ			so	Ꮛ	su	Ꮜ	sv	Ꭱ
da	Ꮝ	ta	Ꮞ			de	Ꮟ	te	Ꮠ	di	Ꮡ	ti	Ꮡ	do	Ꮢ	du	Ꮣ	dv	ᶆ
dla	Ꮤ	tla	Ꮥ			tle	Ꮦ			tli	Ꮧ			tlo	Ꮨ	tlu	Ꮩ	tlv	Ᵽ
tsa	Ꮪ					tse	Ꮫ			tsi	Ꮬ			tso	Ꮭ	tsu	Ꮮ	tsv	ᶜ
wa	Ꮯ					we	Ꮰ			wi	Ꮱ			wo	Ꮲ	wu	Ꮳ	wv	Ꮶ
ya	Ꮴ					ye	Ꮵ			yi	Ꮶ			yo	Ꮷ	yu	Ꮸ	yv	ᵬ

B. Cree[a]

-ê		*-i*		*-o*		*-a*		*Final C*	
ê	ᐁ	i	ᐃ	o	ᐅ	a	ᐊ	-h	ᐦ
pê	ᐯ	pi	ᐱ	po	ᐳ	pa	ᐸ	-p	ᐠ
tê	ᑌ	ti	ᑎ	to	ᑐ	ta	ᑕ	-t	ᐟ
kê	ᑫ	ki	ᑭ	ko	ᑯ	ka	ᑲ	-k	ᑊ
cê	ᒉ	ci	ᒋ	co	ᒍ	ca	ᒐ	-c	ᒼ
mê	ᒣ	mi	ᒥ	mo	ᒧ	ma	ᒪ	-m	ᒽ
nê	ᓄ	ni	ᓇ	no	ᓅ	na	ᓈ	-n	ᐣ
sê	ᓵ	si	ᓷ	so	ᓴ	sa	ᓯ	-s	ᐢ
šê	ᔆ	ši	ᔇ	šo	ᔅ	ša	ᔉ	-š	ᐧ
yê	ᔦ	yi	ᔨ	yo	ᔩ	ya	ᔪ	-y	°/⁺
wê	ᐝᐁ	wi	ᐝᐃ	wo	ᐝᐅ	wa	ᐝᐊ	-w	ₒ
rê	ᕒ	ri	ᕓ	ro	ᕕ	ra	ᕗ	-r	˟
lê	ᓓ	li	ᓕ	lo	ᓗ	la	ᓚ	-l	˟

[a] A dot above the syllable (except for the *-ê* series) marks a long vowel.

2 Writing and Language

The theoretical aspects of writing systems presented here are grounded in a fundamental observation: writing is not like language, and it is not like language for biological reasons. The human language faculty evolved over some

many generations, so that no human infant can avoid learning the language of the environment. No child, however, can learn to read or write simply by watching other people read or write: explicit instruction is required. For writing is so recent (anthropologically speaking) that no special capacity for it can have evolved – especially since literate populations have not reproduced in preference to nonliterate ones!

From this observation it follows that writing need not be structured or described in the same way as language, and in fact some language-derived analytical tools are not so well suited to writing. The linguistic terms phoneme, morpheme, and so on refer to an unconscious property of language (and other realms of human behavior). Each item in a class of "-eme"-designated things is an abstraction, its identity defined by its contrasts with all the other items in that class, and comprising a group of instantiations of the thing. Thus the English phoneme /t/ includes the conditioned allophones [t^h] (in most circumstances), [t] (after /s/), and [ˀ] (sometimes for some speakers); the English morpheme {past} includes the conditioned allomorphs /t/, /d/, and /ɪd/. Every language includes a fairly small inventory of phonemes, and every morpheme is realized with phonemes; every stretch of speech is made up entirely of morphemes, which are made up entirely of phonemes. Here the unconscious-ness is important: since writing is not an unconscious, built-in feature of a mind (as language is), it cannot a priori be assumed to be analyzable in a parallel way. Rather, all writing systems were at some point consciously devised (and, not infrequently, are deliberately modified). The phonemic organization of various phenomena was recognized only a century or so ago, so it is not surprising that the designed writing of language differs in several ways from the evolved speaking of language.

First, writing systems, unlike languages, do not all operate the same. Different writing systems relate to the sound systems they record in fundamentally different ways (in at least the six types identified above). These concern both the amount of speech each symbol represents, and the level of analysis the symbols embody.

Second, despite American structuralist attempts to approach writing as a subsystem of language, writing systems do not work like linguistic systems; there is no "emic" level, and the popular term *grapheme* is misleading. For instance, many alphabets use a pairing of symbols – capitals and lowercase, majuscule and minuscule – that has no equivalent in sound systems. Arguments can be made on both sides of the question as to whether ‹A› and ‹a› are members of the same grapheme (*allographs*). But more basically, no coherent definition of grapheme can be agreed on. Is it (like a phoneme) one of the set of elements comprising a writing system? (Then ‹A› and ‹a› might both be graphemes . . . but how is their relationship to be captured?) Is it (like a tagmeme) a correlation of sound and symbol? (Then ‹ea›, ‹ee›, and ‹e–e› might all be graphemes of English . . . but is ‹ough› then several different graphemes?) Is it (like a morpheme) a minimal extent of something? (Then the Mesopotamian cuneiform signs in figure 3.1 might all contain the same

grapheme . . . though there is no phonetic or semantic similarity among these signs, and the recurring pattern by itself is *xi*.) The difficulty is that all these characterizations are reasonable for different writing systems, but no one characterization fits everything one might be tempted to call a grapheme. The upshot is that *grapheme* has become nothing more than a pretheoretic, fancy, scientific-sounding word for "letter" or "character" and ought not to be part of technical discourse. ("Allograph," however, remains useful for conditioned variants of lettershapes, as in the final variants in Greek and Hebrew, or the conjoined consonants in Indic scripts.)

Third, language is constantly changing, while writing generally obeys tradition and does not readily respond to changes. Simplification in some areas of language is accompanied by complication in other areas, as a language's overall "efficiency" tends to remain constant; but a script's efficiency – its "goodness of fit" to its language – is maximal when it is devised, and deteriorates thereafter.

Fourth, writing systems can be altered by fiat. Kemal Atatürk could not have ordered the minority peoples of Turkey to stop speaking their languages and use only Turkish, but he could decree that the Turkish language would be written with a Roman alphabet rather than an Arabic one beginning on November 3, 1928. Noah Webster could not successfully tell Americans to not split infinitives, say, but he could successfully recommend dropping the ‹u› from words like ‹colour›.

(Fifth, and in the wider picture probably most important, written language differs in significant ways from spoken language; the way most directly related to the physical existence of writing is the evanescence of speech versus the protracted availability of writing. Questions of literacy and society, of literacy and the individual are beyond the scope of this chapter [see Street and Besnier 1994].)

Writing systems, then, must be investigated on their own terms. Their changes in appearance over the centuries – their "outer form" – have attracted the most study and are well documented (see note 2), but most interesting to this author are questions of the origin of writing and the relation of the graphic shapes of script to the phonological shapes of language – their "inner form" (Coulmas 1996: 234).[4]

My approach to the origin of writing arose from dissatisfaction with the received view that there are three types of writing system – logography, syllabary, alphabet – and that the history of writing systems shows that all development has proceeded in that order and can only do so, with alphabets as the last and "best" type. This view is most closely associated with I. J. Gelb, and in order to make it work, he had to claim that what underlay the Greek alphabet was a syllabary. But since the Phoenician script does not explicitly denote syllables, Gelb had to claim that the characters of the Phoenician and other Northwest Semitic scripts in fact recorded syllables – but syllables with indeterminate vowel. As regards the Indic and Ethiopic scripts, which denote syllables but derive from alphabets, he simply threw up his hands (1963: 188).[5]

Two problems are immediately obvious. It is counterintuitive to call Northwest Semitic scripts syllabaries; and anyway, from only one example of the innovation of an alphabet, a general principle can hardly be drawn.

The solution is to recognize that the old tripartite typology is inadequate, and to replace it with the sexpartite one used above. Once abugidas are distinguished from syllabaries,[6] a different historical sequence can be identified, which no longer privileges the alphabet teleologically. Furthermore, this dichotomy also proves useful regarding the modern creation of scripts: scripts invented by persons who cannot read are syllabaries (not abugidas, not abjads, or alphabets). It can also be seen that it is not really the alphabet that represents the great intellectual achievement in linguistic analysis, but the abjad.

2.1 Origin of writing

The key to the history of writing is the primacy of the syllable. Psycholinguists find that people not literate in an alphabetic script are unable to manipulate portions of the speech stream at the level of the segment (Daniels 1988);[7] phonologists increasingly work with levels of analysis other than that of the segment or individual sound (but none seems to have broken entirely with the C's and V's of alphabet-based analysis). The inventions by untutored writing-inventors record syllables. Many nonliterate peoples keep graphic records that perhaps operate on the level of the word. These records do not turn into writing, however (the "reader" cannot determine exactly what sentences the delineator had in mind). Why did the pictographs used in Sumer, China, and Meso-America turn into writing systems? My view is that it is because the Sumerian, Chinese, and ancient Mayan languages were largely monosyllabic (meaning that most morphemes are just one syllable long). Thus each pictograph representing a word also represented a single syllable. It was thus easy, via the rebus principle, to record other similar-sounding words, words that did not lend themselves to pictography since they did not denote simple objects – as in the Sumerian example *ti* "arrow" for *ti* "life." Grammatical morphemes too were soon included in the script stream, and writing was accomplished.

Writing is thus defined as *a system of more or less permanent marks used to represent an utterance in such a way that the utterance can be recovered more or less exactly without the intervention of the utterer*. Implicit in this definition is the insistence that all writing is phonologically based, as stressed by John DeFrancis (1989); excluded are what Gelb calls "forerunners of writing" (DeFrancis shows that none of the "forerunners" actually "foreran" writing) and what Sampson calls "semasiographic systems." In Mesopotamia, what has been identified as an early accounting device, small clay objects ("tokens") that may have served as counters for commodities and were sometimes gathered inside a hollow ball of clay, sometimes after being impressed on the outside of the ball, might have prompted the notion of incising pictographs on lumps of clay and might underlie the shapes of early numerals in cuneiform (cf. Schmandt-Besserat

1992). The suggestion that the shapes of some tokens relate to early, abstract cuneiform signs is purely speculative, since there is no way to know what any particular token may have represented, nor whether there was any sort of uniformity in such representation across the vast extents of time and space from which they have been recovered.

Explanations for the fact that Egyptian hieroglyphics record only consonants are embryonic. But since Egyptian writing never become purely phonetic – logographs and determinatives remained fully in use to the very end of the tradition – we must turn to the abjad for the second great advance in writing, the first that can truly be called an invention. Evidence exists that Mesopotamian scholars recognized an affinity between signs for syllables beginning with the same consonant, and affinity between signs for syllables ending with the same consonant; but there is no evidence that affinity between (what we recognize as) the same consonants at the beginnings and ends of syllables was recognized. (In the Chinese grammatical tradition, syllables were identified according to their initial [consonant] and everything else [vowel + tone + final consonant].) So the greatest stroke of genius in the history of writing was the recognition that syllable-beginnings could be identified with syllable-endings, and the resulting unities could be represented by a single symbol wherever in a word they occurred. These symbols are the consonant letters. And, of course, many fewer consonant letters than syllable signs are needed for just about any language.

2.2 *Script direction*

I suspect, too, that the stroke of genius came from a left-handed individual. Most people are right-handed; most writing runs left-to-right ("dextrograde") or top-to-bottom. These are convenient directions for avoiding smeared ink (or clay) and for clear sight of the line of writing. But the direction of the earliest Northwest Semitic writing (and also the normal direction for Egyptian writing) is right-to-left ("sinistrograde"). This choice makes sense if it was initially made by a left-hander – a left-hander of great prestige, as would certainly befit the inventor of a writing system so much easier to learn than a syllabary or, especially, a logosyllabary (or logoconsonantary). Script direction proves to be a very tenacious attribute of a writing system: so long as the tradition remains unbroken, the direction does not change. Only with a "brusque" transition is an alteration found: this happened with the transfer to Greece: the earliest Greek inscriptions are boustrophedon (running in opposite directions in alternate lines), then the left-to-right order prevails. The transfer to Etruscan must have been early and "gentle," since the direction remained sinistrograde, but the Latin adaptation was less so, since early boustrophedon soon gave way to dextrograde.

Script direction depended on external factors. The Iranian languages all remained "gently" sinistrograde. Syriac scribes, however, would avoid the

mechanical problems in such a script by rotating the page 90° counterclockwise and writing downward and left to right, turning the page back for reading in the traditional direction. (This accounts for the skewed orientation of the Greek-letter vocalizations.) Perhaps this practice was maintained for Sogdian and along with the dominant Chinese culture accounts for the columnar (and left-to-right) writing of the Uyghur, Mongolian, and Manchu traditions. hPags-pa, too, doubtless imitates this tradition. But another local script derived from Tibetan, the Lepcha of Sikkim, was written in columns from right to left – as if Tibetan was rotated clockwise to attain columns in imitation of Chinese.

2.3 Script transmission

Many "gentle" steps brought Aramaic script via Iranian and Altaic languages far to the east in Asia, as described above. But it is the "brusque" transfers that lead to the development of new script types. The accidental alphabet of (Indo-European) Greek has been mentioned. But at the other end of the ancient Near East, the Persian Empire impinged on the Indian world. The Persians brought the Aramaic abjad with them; and some bilingual inscriptions, in Aramaic and (Indo-European) Prakrit, were erected in the northern borderlands. But the Prakrit is not simply written in a variant of Aramaic script; as with Greek, vowel notation was added: not, though, with letters inserted among the consonants, but in abugida style. What lay behind this innovation? The rich grammatical tradition associated with the name of Pāṇini was already well developed by the time writing appeared in India, and it fully understood syllables, vowels, and consonants.

Tibet, too, supported a grammatical tradition. The Tibetan language is typologically quite different from the inflecting Semitic and Indo-European languages met so far: it is isolating, so it could be advantageous for a Tibetan script not to merge adjacent morpheme-final and -initial consonants as was done in the Indic scripts (see figure 3.3). Thus the syllable boundary marker was devised, while the basic abugidic principle of inherent basic vowel plus appendages for other vowels continued in use. Significantly, neither Iranian nor Mongolian culture supported a grammatical tradition, and the hPags-pa script gives up the Tibetan innovation; the morphological type differs yet again, and perhaps explicit syllable boundaries are less important for an agglutinative language like Mongolian. Lastly, as already mentioned, the Korean grammatical tradition played an important part in the design of the most sophisticated script yet devised. Every script reflects some degree of "native speaker analysis" (O'Connor 1983); the lesson of the Asian sequences of transmission is that real innovation in script transfer must be informed by grammatical understanding of the language that is to be written – metalinguistic knowledge of one's language: the result of deep study, not simple copying.

Ꭱ	Ꭰ	Ꮃ	Ꮅ	Ꮆ	Ꮉ	Ꮋ	Ꮕ	Ꮑ	Ꮓ	Ꮖ	Ꮙ	Ꮚ	Ꮜ	Ꮝ	Ꮣ	Ꮤ
e	a	la	tsi	nah	wu	we	li	ne	mo	gi	yi	si	tlv	o	lu	le	ha
Ꮻ	Ꮼ	Ꮽ	Ꮾ	Ꮿ	Ᏹ	Ᏻ	Ᏽ	᏿	ᐁ	ᐃ	ᐅ	ᐍ	ᐟ	ᐡ	ᐣ	ᐥ
wo	tlo	ta	yv	lv	hi	s	yo	hu	go	tsu	mu	se	so	tli	qui	que	sa
ᐧ	ᐩ	ᐪ	ᐬ	ᐭ	ᐯ	ᐱ	ᐳ	ᐵ	ᐷ	ᐹ	ᐻ	ᐽ	ᐿ	ᑁ	ᑃ	ᑅ
qua	no	ka	tsv	sv	ni	ga	do	ge	da	gv	wi	i	u	ye	hv	dv	gu
ᑇ	ᑉ	ᑋ	ᑎ	ᑏ	ᑑ	ᑓ	ᑕ	ᑗ	ᑙ	ᑛ	ᑝ	ᑟ	ᑡ	ᑣ	ᑥ	ᑧ
tso	quo	nu	na	lo	yu	tse	di	wv	du	de	tsa	v	nv	te	ma	su	tlu
ᑫ	ᑭ	ᑲ	ᑴ	ᑵ	ᑶ	ᑸ	ᑼ	ᑽ	ᑿ	ᒁ	ᒃ	ᒄ				
he	ho	mi	tla	ya	wa	ti	tle	na	quu	dla	me	quv				

Figure 3.5 Sequoyah's syllabary order (read left to right)

2.4 Letter order

A property possessed by many writing systems with a limited inventory of signs is a canonical order in which the signs are learned and which becomes an organizing principle for lists of words and for other things as well. Such orders may be arbitrary or motivated; and virtually the only motivated sign-order is phonetic.[8] The Indic scripts, following the native grammatical tradition, placed the vowels (in two groups) before the occlusives (back to front of mouth; within each place of articulation voiceless, voiced, and nasal; for each stop unaspirated and aspirated), which are followed by the continuants (see table 3.7). Modern syllabaries (including current usage for Japanese) are usually presented in a consonant versus vowel grid, with the consonants in some phonetically justified order (see table 3.6) – though the order for Cherokee used in textbooks follows the order of the corresponding consonants in English (see table 3.17A). Some of the syllabaries devised in recent centuries in fact have no standard order; others, including Cherokee, seem to be presented in nothing but the order in which the symbols were devised by the creator (figure 3.5).

This is the best we can say for the familiar order *a*, *b*, *c*, . . . as well. Despite centuries of conjecture – involving lettershape, phonetics, the names of the letters, and doubtless other considerations – no convincing account has ever been suggested. This order is found in the earliest abecedaries, from the fourteenth century BCE – and any speculation must take into account that five letters were dropped from the original 27 (seen in Ugaritic) to give the Hebrew-Aramaic sequence (see table 3.3). This sequence is modified in Arabic to bring together the letters that share a common basic shape and are distinguished

only by dotting. Less familiar is the order recently discovered to have been used for the ancient South Arabian letters (and still more recently found at Ugarit), to which the standard Ethiopic order is similar (see table 3.11). The (North) Semitic order is also known in Ethiopia – since the Hebrew letter names appear as headings to the 22 parts of Psalm 119 (118) in the Septuagint[9] – where, labeled *abugidā*, it serves some liturgical functions. Similarly, when the ancestral order is referred to in Arabic, it is called *abjad*^{un}. (The vocalization of the Ethiopic word reflects the standard order for presenting the seven notated vowels.)

The ancestral Semitic order remains familiar to modern Arabic-speakers because of the "organizing principle" property mentioned above. The sequence of letters, being fixed, could label any sequence of things. This is equivalent to a sequence of ordinal numerals – and even after dedicated characters for numerals (as opposed to tallies: 1 stroke for 1, 2 strokes for 2, etc.) were introduced (first in India in the first half of the first millennium CE, then to the Islamic world around 800, including zero, thence to Europe around 1000), letters have continued to be used as numerals in limited contexts in Greek, Greek-derived, and Semitic scripts. Arabic's letters have been reordered with the additional ones inserted by shape within the inherited sequence, but the inherited numerical values are not altered; the new letters represent the values 500–1000.

Several alphabets have retained letters unneeded for any phonetic value because of already associated numerical values (such as Greek Digamma = 6). This phenomenon and the consistency of correlation of each nonad of letters with an order of numbers (Hebrew Alep–Ṭet = 1–9, Yod–Ṣade = 10–90, Qop–Taw = 100–400) leads Gamkrelidze (1994) to see this as a guiding principle in the creation of the Greek-based Eastern Christian alphabets, which do suspiciously contain multiples of nine letters. If this principle had been in operation from the beginning, however, one would expect five empty-letter numerals in Hebrew-Aramaic script, preserving the earlier total of 27.

Letters added to an abjad or alphabet are usually ordered at the end, as with the Greek "supplementals" after Tau, which corresponds to Taw, the last letter of the parent abjad – and, in fact, as with the last three Ugaritic letters. Sometimes letters are inserted according to graphic similarity (as in Arabic) and sometimes phonetic (as in Cyrillic). Armenian represents an exception, where the framework of the Greek order is discernible but no principle can be found for the placement of the additions.

2.5 *Letter names*

For letters to be learned in an order, they need to have names (table 3.18). Names of letters either are words in the language they record (Northwest Semitic), or they refer in arbitrary patterns to their sound (Greek; Latin). It is not clear which came first – what may be the earliest list of letter names

Table 3.18 Letter names (see also table 3.10)

Ugaritic?	*Hebrew*	*Greek*	*Arabic*	*Ethiopic*	*Armenian*	*Old Slavic*
a	ʾā́leph	alpha	ʾalif	älf	ayb	azъ
be	bēth	bēta	bāʾ	bet	ben	buky
ga	gī́mel	gamma	ǰīm	gäml	gim	vědi/vědě
xa					da	glagoli
di	dā́leth	delta	dāl	dänt	eč‘	dobro
u	hē	e psilon	hāʾ	hoi	za	jestъ/estъ
wa/i/u	wāw	u psilon	wāw	wäwe	ē	živěte
zi	záyin	zēta	zāy	zäi	ət‘	(d)zělo
ku	ḥēth	ēta	ḥāʾ	ḥaut	t‘o	zemlja
ṭi	ṭēth	thēta	ṭāʾ	ṭäit	žē	i, ižei
?	yōdh	iota	yāʾ	yämän	ini	iže
?	kāph	kappa	kāf	kaf	liwn	dervь
?	lā́medh	la(m)bda	lām	läwə	xē	kako
?	mēm	mu	mīm	mai	ca	ljudije
?	nūn	nu	nūn	nähas	ken	myslite
?	sā́mekh	sigma	sīn	sat	ho	našь
?	ʿáyin	o micron	ʿayn	ʿäin	ja	onъ
pu	pē	pi	fāʾ	äf	łat	pokoj
ṣa	ṣādhē		ṣād	ṣädäi	čē	rьci
qu	qōph	(qoppa)	qāf	qaf	men	slovo
ra	rēš	rhō	rāʾ	rəʾəs	yi	tvrdo
śa	śīn/šīn	(san)	šīn	šäut	nu	ukъ/ikъ
xa					ša	frtъ
tu	tāw	tau	tāʾ	täwə	o	cherъ
			θāʾ		č‘a	otъ
			xāʾ	xärm	pē	ci
			ðāl		ǰē	črvь
			ḍād	ḍäppa	ṙa	ša
			ẓāʾ		sē	štja
			ɣayn		vew	jerъ
				ṗäit	tiwn	jery
				psa	rē	jerь
i					c‘o	ětь/jatь
u					hiwn	ju
zu					p‘iwṙ	ja
					k‘ē	(je)
					u	jusъ malyj
					fē	jusъ malyj
						jotirovannyj
						jusъ bolьšij
						jusъ bolьšij
						jotirovannyj
						ksi
						psi
						(thita)/fita
						ižica
						azъ

(incompletely preserved) gives a single syllabic Mesopotamian cuneiform sign opposite each Ugaritic letter. Many of these correspond to the beginnings of the names known much later for Hebrew and Aramaic, but some do not; and it is not easy to imagine why a scribe would not have recorded the letters' full names if they had existed. Most of the Hebrew/Aramaic names are words in Northwest Semitic, a few are not, and their earliest attestation is the aforementioned Septuagint passage. Interpretations of the "Proto-Sinaitic" inscriptions as Semitic based on reading their pictograms according to the initial letters of words for the objects depicted – the *acrophonic principle* – are thus unreliable. Some of the Ethiopic letter names, including the one that licensed the interpretation of the Proto-Sinaitic snake as *n*, are words only in Hebrew, not in Ethiopic, suggesting that the names (not used in Ethiopia) were first assigned by European scholars in the fifteenth or sixteenth century when Classical Ethiopic came to their notice.

The Greek letter names are meaningless in Greek: they are simply borrowed from the Semitic source; apparently the earliest complete list, though, is from ca. 200 CE (Athenaeus 453d), purportedly but dubiously reproducing a fifth-century BCE text. Some of the Arabic names preserve reminiscences of the earlier forms, but the Latin names, which prevail in Europe, are simply phonetic (including CV for stops, VC for continuants), as are those in Georgian and modern Russian, and in the Indic tradition. Words, chosen acrophonically, are used for letter names in Runes, Ogham, Armenian, and Old Slavic.

2.6 *Writing materials*

The shapes of characters can be influenced by the materials on which and with which they are written. We have already seen how cuneiform wedges result from the use of a stylus on clay – where the surface was not conducive to curving lines. Runes are angular because they were scratched into wood; Ogham is straight lines because it was carved on the edges of blocks of stone; and many scripts of India and Southeast Asia are curved because they were incised with a stylus on leaves.

Pigmented liquid (ink, paint) is probably the most common writing medium around the world, applied to surfaces with brushes made from vegetal fibers or hairs, or pens cut from hollow reeds or feathers or forged in steel. The surfaces can be any convenient wall (whether natural as of a cave or cliff, or built), or more portably a clay pot or a potsherd (inscribed sherds are called *ostraca*). The earliest known flexible writing surface is papyrus, prepared from the split pith of a reed native to the Nile valley. Animal skins appear subsequently: leather, prepared by tanning, found from the first millennium BCE, and parchment, somewhat later, prepared by liming. A reusable medium was wooden boards hinged together, their inner surfaces coated with wax, on which Mesopotamian scribes impressed wedges and Greek and Roman scribes scratched letters with a stylus (few of these fragile items have survived, so we cannot be

certain how long they were in use). Paper, which is made from macerated, compacted vegetal fibers, was invented in China early in the Common Era and came west with Muslim contact, eventually superseding the other candidates.

Printing from movable type was devised in East Asia – probably Korea – early in the Common Era, and (perhaps not independently) by Johannes Gutenberg in Mainz in the 1450s. Gutenberg's techniques merged the skills of the goldsmith (for casting type), the vintner (for the press), and the chemist (for the ink). Quick, identical reproductions of texts made possible both religious reformation and the development of science (Eisenstein 1979), but widespread literacy awaited mechanical printing and typesetting in the nineteenth century.

Individual mechanical aids for writers followed: typewriters, cheap pencils, fountain pens, ballpoint and fibertip pens. A feared post-literate society of broadcast media seems now to have been forestalled by the worldwide network of personal computers, on which international communication is again achieved in writing.

3 Writing and Scholarship

Writing is indispensable for civilization – but entirely irrelevant for language. Most of the thousands of human languages were never written until recent years, and their speakers were none the worse for it. Their cultures were full and rich, lacking only accountancy and science. Everything else that is written need not be: poetry, narrative, law, and their apotheosis, scripture, are all part of every oral culture. Only in a city is the community so large that letters must be sent to communicate personal messages – and only when records of commerce can be kept can a city *be*. Cities are where production does not link directly with consumption: farmers and ranchers provide food, artisans provide goods, builders provide shelter, and administrators coordinate their exchange. Without writing, there is no administration.[10]

But cities characterize only a handful of human societies, and the vast majority of human languages never had written forms of their own. The discovery that languages other than the classical ones were every bit as rich as Greek, Sanskrit, and Chinese – a discovery due largely to the investigation of Native American languages by scholars originally trained as Indo-European philologists, on the whole – led linguists to concentrate on unwritten languages and then to devalue the study of written records in favor of fieldwork. Recently a reaction (associated with the "Toronto school" of literacy studies) to this view has set in, which in its most extreme form claims that there was no true literacy before the Greek alphabet, even in the ancient Near East: that the alphabet itself is necessary for elevated linguistic expression. What this attitude reveals is little more than ignorance of both the literary record of non-alphabetic societies (which is generally known to the partisans only through

translation) and of the poetic accomplishments of nonliterate societies (represented most familiarly, of course, in the supposed foundational work of western literacy, the Homeric epics).

With that *parti pris*, we may turn to the branches of scholarship that have studied writing systems.

3.1 Philology, epigraphy, and paleography

Philology is the study of texts in the broadest sense. The preliminary task of philologists includes recovering and establishing the documents themselves, determining the orthography, grammar, and lexicon of their language, and reconstructing their history and context. Then their real work begins: interpreting the texts and the entire culture that underlies them. Among the subdisciplines of philology are *epigraphy*, *paleography*, and *diplomatics* (the study of documents).

A distinction is often made between writing incised on solid surfaces and writing applied with a tool to flexible surfaces. The former is the province of epigraphy, the latter of paleography. A goal of both fields is tracing minute variations in *ductus* – the complex of features characteristic of a single scribal community – from generation to generation, from atelier to atelier, which might enable the dating of a text that has no explicit indication of when and where it was written, such as a colophon. This has largely been pursued as a purely descriptive study, with little attention to the physical processes of writing – movements of hand and fingers that always want to expend less effort, in competition with the cognitive need to keep characters recognizably distinct. In this tension lies legibility.

The appearance of a script is also closely connected to the prevailing esthetic of its society, as is familiarly seen in the affinity of spiky German hands to the pointed arches of Gothic architecture, and of curvaceous Italian hands to the rounded arches of Romanesque. Worldwide, artistic approaches to writing, calligraphy, mirror the arts of the society, sometimes even becoming the dominant decorative art, as in much of the Islamic world.

The more routine task of epigraphers and paleographers is the compilation of corpora of inscriptions and texts: the raw materials from which philologists extract descriptions of cultures, and linguists extract descriptions of languages and language change. Often, the preparation of a corpus of unreadable texts (Linear A, Indus Valley script) is all that can be done until some genius can discover the hint that makes it possible to read them.[11]

3.2 Decipherment

Ancient and mysterious scripts captured the imagination of adventurers whenever they came upon them, but not until the middle of the eighteenth century

did anyone succeed in reading one whose interpretation had been forgotten with the culture that created it. The script that received this honor was not a specially worthy one; it was the Palmyrene variety of the cursive Aramaic group found throughout the Near East at the beginning of the Common Era. The rulers of Palmyra would place inscriptions in both Greek and Palmyrene Aramaic on the monumental columns that lined the public spaces, and in 1756, accurate copies of several such pairs were published in London and Paris as engravings. Virtually overnight, the abbé Jean-Jacques Barthélemy was able to interpret the Palmyrene. His method exemplified many of the principles that have been used many times since: identify a bilingual text; locate proper names; compare known scripts; guess what language is represented; determine from the number of different characters the likely type of script. (In short order, Barthélemy also deciphered Phoenician and Imperial Aramaic.)

Prior to all the steps in the actual decipherment, however, and so obvious that it is often overlooked, is the necessity of accurate reproductions of inscriptions in the unknown script. For a century and a half, photography has been available, but many important decipherments were accomplished in the century before that. Before Barthélemy, there had been publications of Palmyrene inscriptions going all the way back to 1616, none of them amenable to decipherment (yet no one who had not visited the inscriptions *in situ* could know that)! Fortunately, for both the best-known decipherment and also for arguably the most important one, the publications available to the decipherers were of the highest quality.

The most familiar deciphered script, of course, is Egyptian hieroglyphs. Napoleon invaded Egypt with an army of scholars as well as an army of soldiers. Over several years, the scholars prepared painstaking representations of the wondrous antiquities of the Nile, including scores of inscriptions on both monuments and papyri. Among the inscriptions was a large slab found in 1799 (and forthwith seized as booty when the British gained an advantage, so that the Rosetta Stone has been housed in the British Museum since 1802) inscribed in Greek and in demotic and hieroglyphic Egyptian (the hieroglyphic portion largely broken away). It was immediately seen that this could be the key to interpreting the Egyptian inscriptions that had fascinated Europeans since Classical antiquity – but the key could not be turned so long as the mental machinery was mired in the millennial mirage that the hieroglyphs were "ideograms" or mystical, occult symbols. An English dilettante (or polymath), Thomas Young, identified corresponding passages in the Greek and demotic texts but neglected the incomplete hieroglyphic version.

Meanwhile Jean-François Champollion, a young man from Grenoble, had resolved to understand ancient Egypt, and he believed the clue lay in the Coptic language, still used in the liturgy of the Christians of Egypt. Around 1820 he noticed that the pharaonic name "Ptolemy" occurred several times in

the Greek in positions corresponding to cartouches (oval frames containing signs) in the hieroglyphs, confirming a long-ago suggestion of Barthélemy's that they would enclose royal names; but since that was the only royal name preserved in the hieroglyphic, he could not try assigning the phonetic values he suspected the Egyptians would have used to write foreign names (here, Greek). Fortunately, the name of Cleopatra appeared on an obelisk that had been in England since 1813 (Champollion may have known it from Young's publication), and there is sufficient overlap in the names that he could pair signs with sounds. Other names in Greek and Latin gave him several other phonetic values.

Champollion's true breakthrough came when he noticed a cartouche containing an obvious "sun" logogram followed by an unknown sign and two *s*'s. "Sun" in Coptic is *rē*, and the decipherer, against all expectations, guessed that the name was Ramses – showing that Egyptian names, too, could be written phonetically. This gave him the courage to search for Coptic words in the Rosetta prose, and soon he could read Egyptian.

The most important decipherment recovered Mesopotamian cuneiform. The basic materials here came from a brief span of ancient history, the Persian Empire. From the late sixth to the mid-fourth centuries BCE, kings Darius, Xerxes, and their successors inscribed on the walls of their constructions, their monuments, and on a cliff at Behistun, propagandistic annals and dedications in three cuneiform scripts. The most prominent was the simplest, comprising a few dozen different characters, the other two considerably more complicated. The prominence suggested to a junior faculty member in Göttingen, Georg Friedrich Grotefend, that the simplest script represented the rulers' own language, Persian. On the basis of Antoine Isaac Sylvestre de Sacy's recent decipherment of some Sassanian inscriptions (representing an Iranian empire a few centuries later), he expected to find introductory expressions along the lines of "X, great king, son of Y, great king." The names of the Persian kings were known, in Greek guise, from the Classical historians. Sure enough, Grotefend found the repetitious pattern, plausibly interpretable as "Xerxes, great king, son of Darius, great king, son of Hystaspes" – who was not a king. His discovery was announced in 1802, and over the next several years, scholars were able to clarify the characteristics of Old Persian.

Note that the initial breakthrough did not involve a bilingual; it was achieved through the insight that names known elsewhere could be expected in the unknown text. Such a correspondence can be called a *virtual bilingual*. The names in the Persian trilinguals did provide the initial clue to the other two languages, but they were soon superseded by a wealth of inscriptional material that became available during the first decades of the nineteenth century. Edward Hincks, an Irish clergyman, applied himself first to the trilinguals (coming up with an initial list of values for the signs of the second script in 1846), and then turned to annalistic materials coming from Babylonia; he used Semitic grammatical patterning to locate signs involving constant root consonants and affixes. His most useful source, though, proved to be a massive

annalistic inscription in yet a fourth language, Urartian, where repetitious formulae provided spelling variants permitting the identification of the vowels of many syllables. By 1852, Hincks had succeeded in reading the third script, the language now called Akkadian, and moreover had identified the first of thousands of fragments of ancient dictionaries that made the study of Akkadian (and Sumerian) something other than decipherment. Meanwhile H. C. Rawlinson had, with great effort, made a copy of the huge, virtually inaccessible Behistun inscription of Darius. This accomplishment, plus his edition of parts of the Persian and Akkadian versions, have generally gotten him the credit for deciphering cuneiform, but he was kept abreast of Hincks's findings and incorporated them into his own work – and Behistun was not published until the decipherment was virtually completed, and had little or no impact.

A number of other decipherments have followed (and a few challenges remain); the most celebrated recent one was Michael Ventris's of Linear B, which proved to record an early form of Greek; here the virtual bilingual was the coincidence of the names of findspots of documents with what seemed to be placenames in the texts: sign values assigned on their basis and plugged into a C?V? grid established by Alice Kober and Ventris revealed familiar-looking inflections. Something similar played a part in the decipherment of Maya glyphs: after Yuri Knorosov interpreted a sixteenth-century Maya-Spanish "alphabet" as a syllabary, and saw in the images pictures that could be named with suitable modern Mayan words, Heinrich Berlin found that distinctive signs were associated with specific places. In both cases the attempt to fit a familiar language (Classical Greek, forms of modern Mayan) to the ancient writings provided the final, if surprising, success. The mysterious script most apt to be deciphered some day is the Indus Valley script used between about 2500 and 1900 BCE. This is likely to prove a fourth independent invention of writing, and the Dravidian family is the likeliest candidate to provide its language.

3.3 Writing, linguistics, and semiotics

Is the study of writing – *grammatology*, as Gelb dubbed it – to be seen as a part of linguistics? The study of written language certainly is. But the fundamental difference between language and writing suggests that perhaps writing is outside the scope of linguistics, especially when linguistics is seen as a part of psychology. Perhaps the study of writing truly belongs as a sister science under the umbrella of semiotics, the study of meaningful systems embracing but transcending language. Semiotic approaches to writing, however, have tended to slight philological concerns, to skip right over the details in favor of ungrounded theorizing. Perhaps when writing systems come back into the ken of linguists, the situation will improve.

The traditional arrangement of Japanese hiragana (each character is used once, to spell out the following poem, attributed to the Buddhist monk Kūkai):	いろはにほへとちりぬるを わかよたれそつねならむう ゐのおくやまけふこえてあ さきゆめみしゑひもせす *Iro wa nioedo chirinuru* *wo waga yo tare zo tsune* *naran ui no okuyama* *kyō koete asaki yume* *miji ei mo sezu*	"The colorful [flowers] are fragrant, but they must fall. Who in this world can live forever? Today cross over the deep mountains of life's illusions and there will be no more shallow dreaming, no more drunkenness."

The Javanese order: ꦲꦤ ꦕꦫꦏ ꦢꦠ ꦱꦮꦭ ꦥꦢ ꦗꦪꦚ ꦩꦒ ꦧꦛꦔ
Hana caraka, data sawala, padha jayanya, maga bathanga
"There were two emissaries, they began to fight, their valor was equal, they both fell dead."

Figure 3.6 "Motivated" canonical orders of scripts

NOTES

I am grateful for the comments and suggestions of the volume editors, and of Jerrold S. Cooper, John DeFrancis, Victor Mair, M. O'Connor, and P. Oktor Skjærvø. Space limitations preclude incorporating them all, which can only be to the detriment of the chapter.

1 Sampson 1985, Coulmas 1989, DeFrancis 1989.
2 Taylor 1883, Cohen 1958, Février 1948/59, Friedrich 1966, Diringer 1948/68, Jensen 1935/69, and Senner 1989. On a smaller scale, but useful, are Nakanishi 1980 and especially Woodard 1996.
3 A few recent items not included there are added in the references here.
4 Coulmas 1996 contains numerous insightful articles on societal aspects of writing. Unfortunately the work is arguably unreliable as to factual matters.
5 A more nuanced statement appears in the revision of this passage in Gelb 1974: 1038.
6 The name "abugida" (borrowed from Ethiopic languages) is used in preference to existing terms like "alphasyllabary," "neosyllabary," "pseudo-alphabet," and "syllabically organized alphabet" in order to stress the independence from both "syllabary" and "alphabet."
7 Prakash et al. 1993 find that even abugidic literacy does not suffice for segmental awareness.
8 The exceptions are Japanese, where the classical arrangement of the 50 characters of the syllabaries spells out a poem, and Javanese, where the 20 C*a* letters spell out a sentence summarizing an etiological tale (figure 3.6).
9 The Septuagint is the Greek translation of the Hebrew Bible, dating from ca. the second century BCE, which underlies all the ancient versions of the Old Testament.
10 The Inca quipu (elaborate knotted cords recording numerical information) fulfilled this function in Andean civilization; noteworthily, their Quechua language cannot be described as monosyllabic.
11 Pope 1975 is the best history of decipherments except cuneiform (for which see *WWS* 145–7 summarizing Daniels 1994).