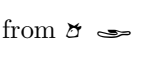
Philology

The alphabet tree

Peter Wilson

1 Introduction

We got from  via  and  to A K N, and  and  in only about 2000 years. This is a short story of how that happened and how we know what the strange symbols mean, e.g.,  and  . The fonts in all the examples were designed for use with TEX, especially by humanities scholars, and are freely available.

1.1 Writing systems

The earliest known writing is from Sumeria dating from about 3400 BC. In some other places the earliest writings discovered date from: Egypt in 3000 BC, the Indus Valley in 2500 BC, Crete in 1900 BC, China in 1200 BC and Central America in 600 BC. These all looked very different.

It seems that initially writing was used for bureaucratic purposes — keeping accounts, recording goods and so on — and a limited writing system was sufficient for this. Later, to record the great deeds of the rulers and especially their names, to promulgate their laws, and to meet the needs of the religious establishment, writing would be extended to a *full writing system*: a system of graphic symbols, or *glyphs*, that could be used to convey any idea. At that point literature became a possibility. All writing systems represent speech in one form or another. Some glyphs represent sounds while others are *semantic* signs that represent either words or concepts; these are called *logograms*.[1]

It appears that early writing systems followed the same general progression. The first actual writing was *pictographic* or *iconographic* where a simple picture designated a real object — a drawing of a deer represented a real deer, for example. Generally the pictures were very simple and abstractions of what we might think of as a drawing. A stylised picture is called a *pictogram*.

Gradually the pictures were formalized and also began to be used to represent relationships and ideas as well as objects. This is called *ideographic* writing. For example, a picture of the moon could represent the idea of night or darkness as well as that of the moon itself. A symbol standing for an idea is a semantic sign and is called an *ideogram*.

A major intellectual step was the invention of the *rebus device*. This is where the sounds corresponding to pictograms are combined to form a word. We have all come across these, often as children's puzzles. For example, a picture of a bee plus a picture of a tray can represent the word 'betray', or more obscurely a picture of a bee plus a picure of a female deer can stand for the word 'behind'. Even a single pictogram can suffice; for instance a pictogram of the sun can be used for both the words 'sun' and 'son'. Consequently symbols can be created that just stand for sounds and then they can be combined to form words. This can markedly reduce the number of symbols required for a full writing system. Symbols representing sounds are called *phonograms*. *Phonetic* writing requires few glyphs. In these writing systems, the glyphs represent sounds. All writing systems are a combination of phonetic and logographic elements but the proportions of these two elements vary among languages. Ideographic scripts essentially have one glyph for each word, and this usually represents the meaning of the word, not its pronunciation. The arabic numerals 1, 2, ... are pronounced in English as one, two, ... but in German as eins, zwei, ... even though the meanings are identical. Mixed systems are where some signs are ideographic and others are phonetic. For example, in English 1st, 2nd, ..., are ideograms with a phonetic component so that we read them as first, second, ... instead of onest, twond, and so on.

Although it does have some phonetics the Chinese script is principally logographic; this may be because of the Chinese language itself. Spoken Chinese consists almost entirely of one-syllable words and there is a limit on the number of short sounds that the human voice can make; the Chinese use something like 400–900 sounds. Many sounds, therefore, have shared meanings — *homophones*. In spoken English the meanings are deduced from the contex and in writing by their spelling. For example: 'Pare me a pair of pears'. In spoken Chinese homophones are partly distinguished by using four levels of pitch (thus increasing the number of different word sounds to some 3000), and by context. In writing there is little possibility of reducing the number of glyphs required to represent the vocabulary from that of an ideographic script. Chinese and Japanese use the same ideographic script although their languages are very different.

At the other end of the spectrum, the Finnish and French systems are much closer to pure phonetics, with some logographs; that is, the phonetic values of the set of glyphs is closely matched to the

[1] From the Greek *logos* = word.

sound of the spoken language. In the English writing system the alphabetic glyphs, a–z and A–Z, represent sounds, the '?' mark represents an idea and glyphs like '$' represent whole words (which can also be spelled out as 'dollar').

From the several thousand characters that an educated Chinese or Japanese needs to know, the English speaker only needs to know 26 (52 if you include both uppercase and lowercase). Monolingual Chinese and Japanese can read and understand each other's scripts because they are based on the same ideographic system, even if they can't understand each other when speaking. Monolingual French and English for example, with their alphabetical writing systems, are not in this happy situation; they can read each other's scripts but with no understanding of either the written or spoken words.

Phonetic scripts can be roughly classified into two major kinds.

- In a *syllabic* system a glyph represents a syllable, usually a consonant followed by a vowel (CV) but can be a VC pair or a consonant vowel consonant (CVC) triple.
- In an alphabetic system a glyph represents either a vowel or a consonant, again with two subdivisions. In some alphabetic systems, like Hebrew, only the consonants are denoted, whereas in a full alphabetic system, like French, both consonants and vowels are fully represented.

By about 3200 BC the Sumerians were using a cuneiform (wedge-shaped) script. They lived in the Fertile Crescent in the area of the Tigris and Euphrates in what is now the Middle East. The Sumerians had over 2000 ideograms. Following their discovery of the rebus device they eventually reduced the number of glyphs in their script to about 600, which included both semantic glyphs and phonograms. The Sumerians wrote on clay tablets using a stylus to impress the marks. Drawing on clay is not easy, which must have given them an impetus to move away from pictograms towards ideograms and then on to phonograms.

The Egyptian hieroglyphic[2] writing system developed roughly in parallel with the Sumerian system. The Egyptians, though, wrote on papyrus with a reed brush or pen, or painted on the walls of tombs. Drawing with these implements is much easier than scratching pictures in clay and thus they did not have such a great need to move towards phonograms. For everyday purposes they did develop more efficient writing methods with their cursive hieratic and demotic scripts. The hieratic script was invented soon after hieroglyphs and was initially the Egyptian everyday business script. The demotic[3] script came much later, around 650 BC, and was then used as the everyday script. The priests, though, continued to use the hieratic script.

Unlike hieroglyphics, which failed to spead beyond Egypt, cuneiform writing became popular and was taken over by the Babylonians when they conquered the Sumerians in 1720 BC.

There is evidence that there was a writing system in Crete about 3000 BC. A thousand years later a syllabic script called Linear A was in use, and by about 1800 BC this had been replaced by the Linear B syllabic script which was used for writing Mycenean Greek.

By about 1500 BC, Egyptian hieroglyphs included 'alphabetic' glyphs alongside all the others. The Ugaritic alphabetic cuneiform script dates from about 1300 BC.

1.2 The alphabet

Although it may never be possible to describe accurately the origins of the alphabet, scholars are generally agreed that most of the world's alphabets are descended from one that was probably invented about 1600 BC in the Middle East. This was a Semitic alphabet, where Semitic refers to a linguistic family that ranged between the Sinai in the south, along the Mediterranean coast, north through Asia Minor and east to the Euphrates valley; the Canaanites and Phoenicians, among many others, spoke a Semitic language.

As people travelled, particularly as conquerors or merchants, the original alphabet was disseminated geographically and gave rise to several alphabetic branches. Roughly speaking, in the time period 1400–1100 BC, these were Archaic Greek; Old Hebrew which led to Samaritan; Phoenician; and South Arabic scripts which in turn led to Amharic, via Ethiopic, and other obscure scripts like Thamudic and Lihyanic. With only minor exceptions these scripts were used for writing languages from the Indo-European family. All are interesting in their own right, but the most relevant script for us is the Phoenician which is the direct ancestor of not only our modern Latin alphabet but also of many other alphabets and scripts in use today. Figure 1 shows the main descendants of the Phoenician alphabet.

At the earliest time there was no fixed direction to writing. It could be left to right or right to left, randomly, or at times lines would alternate between left to right and then right to left. This alternation

[2] From the Greek for sacred engraved writing.

[3] From the Greek 'demotikos' meaning 'in common use'.

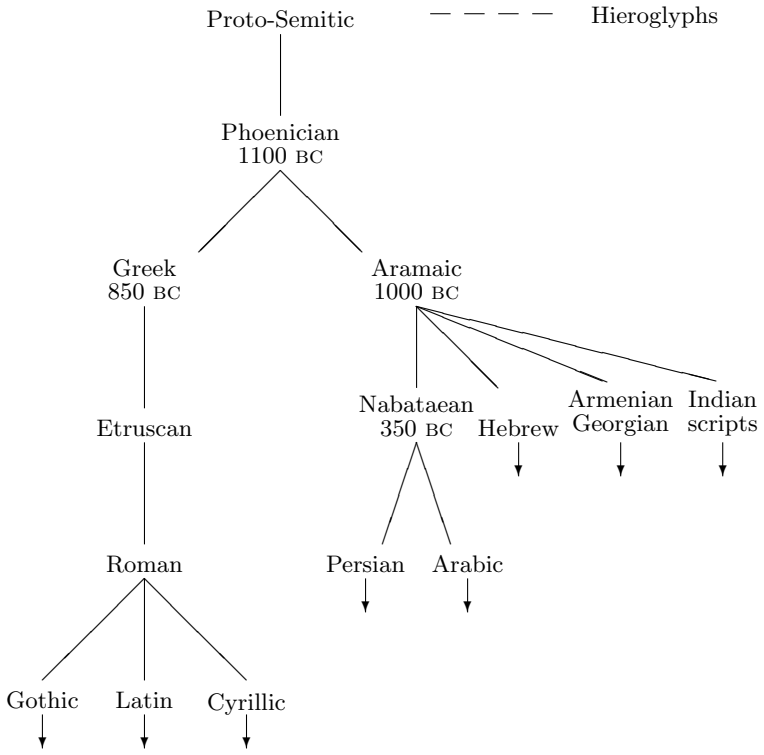


Figure 1: The alphabet tree

of writing direction is termed *boustrophedon*[4] writing. Typically, in boustrophedon writing, two sets of glyphs were used, one being the mirror image of the other, so those who were literate would have twice as many characters to remember.

2 Changes and decipherment

As we will see, scripts are remarkably resistant to change, but nevertheless they do. There are two main reasons why a script should change.

One is a change in technology, which during the period considered comes down to the writing materials used. These ranged from using pointed implements to make marks in soft clay, reed brushes to paint or write on a smooth surface, and scratching or chiseling letters into hard stone.

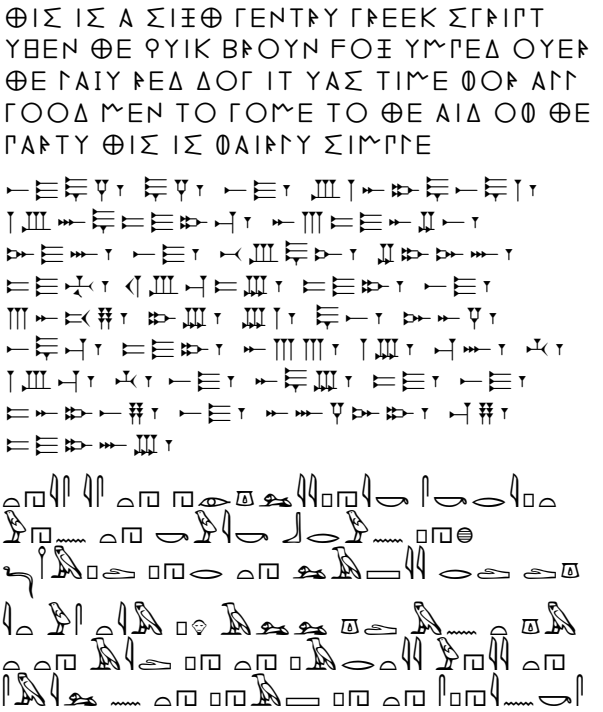
The other reason is a script starting to be used to denote a language that it was not designed for. Every script represents the sounds of a language, and not every language has the same set of sounds. It is easy enough to drop characters that do not represent a sound in a language but it seems much harder to introduce new characters for new sounds. If there are redundant characters they are given new sound values; only later may new characters be introduced. However, it is more likely that existing characters will be decorated to denote new sounds. For example, French and English use the same set of alphabetic characters, but the French add accents to some characters (for instance, è, é, ...) to denote sounds used in spoken French but not in English.

Thus, a particular script may be used for several languages, and over a period of time one language may be captured using several scripts. As the ages pass civilisations also pass, and their scripts may be forgotten until rediscovered by archaeologists, by which time their languages may also have been lost. According to Robinson [Rob02, p. 262] decipherment 'is a process of deducing from texts a known or plausibly reconstructed language that accounts for the patterns of sign use in texts.' A decipherer is faced with three possibilities:

- A known script, in the sense that the 'meanings' of the glyphs are understood, and an unknown language;
- an unknown script and a known language; or, worst of all,
- both the script and language are unknown.

[4] From the Greek for 'as an ox plows a field'.

A few very brief descriptions of some decipherments are given later. In case you would like to understand better some of the problems of decipherment, you can try your hand at the following.



Just to help you along, the scripts are all written left to right, even if the original scripts were written in different directions, and are more or less transliterations of the following English paragraph.

> This is ——. When the quick brown fox jumped over the lazy dog it was time for all good men to come to the aid of the party. ——.

In real life the task is much harder than the examples imply. A representative set of example texts has to be gathered and then two things have to be done before the hard work starts — determine the writing direction, and determine the kind of script.

The writing direction is often determinable by noting whether the text is 'set' ragged-right (left to right direction) or ragged-left (right to left direction). Of course some texts may be 'centered' or 'justified', which is not much assistance.

The other preliminary task is to decide on the kind of script. Scripts can be classified into one of three types: phonetic, ideographic, or mixed. In a phonetic script the glyphs represent the sounds of the language. Alphabetic scripts aim at the ideal of one sign per sound, but this ideal is rarely met. Syllabic scripts use one sign for each syllable. The kind of script can usually be determined by counting the number of different glyphs. An alphabetic script typically has between 20 and 30 characters, a syllabary has roughly 30 to 60 characters and more than that indicates ideographic elements. However, if you apply that naively to English writing then it looks like a syllabary as there are 52 alphabetic characters (26 in each case) plus numerals and punctuation. The same problem may arise with ancient scripts, as it is not always easy to decide whether two similar looking characters are the same or not. Table 1 from [Coe99] lists the numbers of glyphs in some different kinds of scripts.

There is an empirical formula for estimating the probable number of signs in a script from a small sample of the script [Rob02, p. 310]. It seems to work for modern languages and scripts such as Arabic, English, and Japanese *kana* as well as ancient ones such as Linear B.

In a small sample of an alphabetic or syllabic writing system consisting of a total of L characters of K different kinds then the probable number of symbols S forming the alphabet or syllabary is, subject to various restrictions not enumerated here, given approximately by the formula

$$S = L^2/(L-K) - L. \qquad (1)$$

Applying this formula to the previous paragraph where, ignoring the the formula itself, uppercase

Table 1: Numbers of individual glyphs in writing systems

Logographic		Syllabic		Alphabetical	
Sumerian	600(+)	Persian	40	English	26
Egyptian	800	Linear B	87	Anglo-Saxon	31
Hittite	497	Cypriot	65	Sanskrit	35
Chinese	5,000(+)	Cherokee	85	Hebrew	22

Table 2: Common transliterations and their pronunciation

Symbol	Pronunciation
d	d, as in 'did'
$\underline{d}$	dj, like j in 'joke' or di in the French 'dieu'
g	hard g, as in 'get'
h	h, as in 'home'
$\dot{h}$	an emphatic h, sounded in the throat
$\underset{\smile}{h}$	ch, as in the Scots 'loch'
$\underline{h}$	softer than $\underset{\smile}{h}$, like the ch in German 'ich'
$\hat{\imath}$	y, as in 'yea'
k	k, as in 'kit'
$\dot{k}$	k in the back of the throat, like the Arabic q in *Qur'ân* (Koran)
r	a trilled r, as in Scots 'rain'
s	s, as in 'soap'
$\check{s}$	sh, as in 'ship'
t	t, as in 'tub'
$\dot{t}$	t, as in 'tune'
$\underline{t}$	tj
w	w, as in 'wet'
y	y, as in 'yes'
ʒ or '	glottal stop, like the break in the Cockney pronunciation of 'bottle' as 'bo'el' or the American pronunciation of 'Seattle' as 'Sea'el'
'	gutteral, the Semitic ayin

characters and punctuation, there are 232 characters of 24 different kinds (q and z are not used) we get the approximate value of

$$S = 232^2/(232 - 24) - 232 = 26.77$$

for the number of signs in the lowercase English alphabet compared to the actual value of 26 signs.

3 The earliest scripts

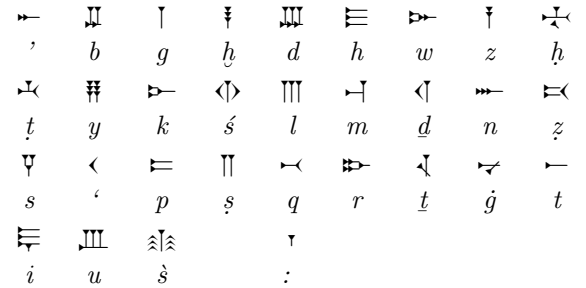
We now show a variety of scripts dating from before 1000 BC, some of which are related. Transliterations into modern Western characters are also given; Table 2 lists the main transliterations used and their pronunciation.

3.1 Sumerian and Ugaritic

The earliest script so far discovered is Sumerian cuneiform dating from about 3200 BC, which had developed from earlier pictograms. At its most bloated it included over 2000 glyphs but as it proceeded through the normal evolutionary process the number of glyphs dropped to about 600 in its final form.

The Ugaritic cuneiform script dates from about 1300 BC and was alphabetical, although like most scripts for the Semitic languages did not include vowels. The script consisted of 30 letters and a ideographic word divider (a short vertical wedge). It appeared to have got the alphabetical idea from contemporary linear scripts. It was used to write a language related to Hebrew. In addition to the typical administrative texts there are a number of mythological texts about the god Baal, which give scholars another view on some Biblical stories.

The order of the glyphs in an alphabetic script is usually revealed by writing found from scribal schools where the pupils were practising their abecedaries. In the original order, which is reasonably typical of Semitic scripts and with the word divider (𐎟) being the last glyph, the Ugaritic alphabet and the modern transliteration is:



The last recorded use of a cuneiform script was in 75 AD, so cuneiform vies with hieroglyphs for the longest period of use of any script.

3.2 Hieroglyphs

Hieroglyphs were used by the Egyptians from about 3000 BC to 400 AD. The script is a mixture of a set of consonantal glyphs, a syllabary, and logograms.

Table 3: A hieroglyph sampler

Glyph	Sound	Meaning	Glyph	Sound	Meaning
	ꜥ	arm		ꜣ	vulture
	b			*b*	leg
	ḥ	ball of string?			
	ḏ	cobra		*d*	hand
	šꜣ	pool with flowers		*ỉr*	eye
				f	horned viper
				*g*	jar stand
	ḫ	twisted wick		*h*	
		eat, drink, speak		*ỉ*	reed
				*pr*	house, building
	ḳ	slope of hill		*k*	basket with handle
	ỉw	walk, run		*l*	lion
	m	pair of ribs?		*m*	owl
				n	water
		palm of hand		*p*	
	tp	head		*ḥr*	face
	wr	small, bad, weak		*r*	mouth
	s	door bolt		*s*	folded cloth
	ṯ	tethering rope		*t*	
	w			*w*	quail chick
	ḫ	ox			door
		rejoice		*y*	pair of reeds
				š	stone, pool
	ỉmy	crossed planks		*awt*	shepherd's crook
	ḳmꜣ	throw stick			

There are approximately 6000 known different hieroglyphs, but fewer than 1000 were in use at any one time.

A short sample of hieroglyphs is shown in Table 3. As an example of Egyptian writing, the following hieroglyphs:

are transliterated as:

wḏ ḥm.f ḥr wrryt.f nt ḏꜥm ỉb.f ꜣw

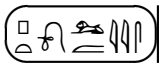
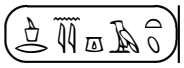
and can be translated as:

His Majesty departed upon his chariot of electrum, his heart joyful.

There were also hieroglyphs for numerals, some examples being: (1), (2), (10), (100), (1,000), ... (1,000,000).

The breakthrough in the decipherment of hieroglyphics came after the Rosetta Stone was discovered in July 1799 near Rashid, which was the ancient Egyptian town called Rosetta, by French soldiers in Napoleon's invading army. The stone carries an inscription in three different scripts: hieroglyphs at the top, which was badly damaged with about half missing; Egyptian demotic script in the middle; and Greek at the bottom. There are 54 lines of Greek with the right hand ends of the last half being damaged or missing. The demotic portion has 32 lines, written right to left, and the right hand ends of the first 14 are damaged. The first half of the lines of hieroglyphs are completely missing and the existing 14 lines, which correspond to the last 28 lines of Greek, are damaged at both ends. The Rosetta Stone is now kept at the British Museum.

The first attempts at decipherment focused on the demotic script. Initial partial decipherments were accomplished by the Frenchman Sylvestre de Sacy (1758–1832) and the Swedish diplomat Johan Åkerblad (1763–1819). The basis was being able to

Hieroglyph	Young's value	Hieroglyph	Young's value
	p		*bir*
	t		*e*
	not essential		*n*
	lo or *ole*		*i*
	ma or *m*		superfluous
	i		*ke* or *ken*
	osh or *os*		
			feminine termination

Figure 2: Young's decipherment of the Ptolemy and Berenice cartouches

identify corresponding names, such as Alexander, Ptolemy, and Berenice, in the Greek and demotic texts. This gave sounds for some of the demotic signs and from this it was possible to show that the script had phonetic components. From the identified signs it was possible to identify some other words such as temple and love. Unfortunately, Åkerblad was convinced that the script was entirely phonetic, which blocked any further progress on his part.

The English polymath Thomas Young[5] (1773–1829) then took up the challenge in 1814. Young was a prodigy; he could read fluently before he was three, and by the time he was fourteen he had studied Arabic, Chaldean, Ethiopic, French, Greek, Hebrew, Italian, Latin, Persian, Samaritan, Syriac, and Turkish. He proved that the demotic and hieroglyphic scripts were not completely distinct and that the Egyptians used a mixed writing system. He was able to decipher much more of the demotic and established the equivalence of many demotic and hieroglyph signs. He determined that the only royal name appearing in the hieroglyph section was Ptolemy. This was spelt phonetically in demotic and he surmised that it was also spelt phonetically in hieroglyphs, corresponding to the Greek (Ptolemaios). Young produced the list of values given in Figure 2.

From an inscription at the temple of Karnak he also had the name of the queen Berenice (Greek Birenike) and for this he constructed the further correspondences, also shown in Figure 2.

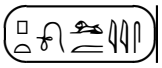
This is about as far as he could get, as he believed that the vast majority of hieroglyphs were ideographic, and phonetic spelling was limited to the names of foreigners. As his mind was so quick he was probably also bored by the time he got to this stage.

The final decipherment was achieved by the Frenchman Jean-François Champollion (1790–1832), who was more open minded than his predecessors. At age ten, on having been shown hieroglyphs by the French mathematician Jean-Baptiste Fourier and being told that nobody could read the strange writing, he decided that he would solve the mystery. To equip himself for the task he studied Arabic, Chaldean, Chinese, Coptic, Ethiopic, Greek, Hebrew, Latin, Pehlevi, Persian, Sanskrit, Syrian, and Zend. By 1822, through systematic analysis of the available material, he showed that the hieroglyphic script had phonetic principles. To progress further he needed to have two or more known names with some hieroglyphs in common so that they could act as a check on any proposed decipherment.

In 1819 W. J. Bankes had had an obelisk moved from Egypt to his home at Kingston Lacy in Dorset, England. The hieroglyphs included two different cartouches and the Greek inscription at the base of the obelisk mentioned Ptolemy and Cleopatra. He noticed that one of the cartouches was identical to that deciphered as Ptolemy by Young, and surmised that the other corresponded to Cleopatra. Bankes had lithographs made of the inscriptions, annotated with his idea about the cartouches, and distributed them in 1821. When Champollion received a copy he made the decipherment shown in Figure 3.

There was a remarkable degree of similarity between the values from the two names, except for the and signs which he explained as being homophones — they could each represent the same sound (*t* in this case).

[5] If you have taken any science courses you have probably heard of Young's Modulus and Young's Rings.




Hieroglyph	Champollion's value	Hieroglyph	Champollion's value
	p		*c*
	*t*		*l*
	o		*e*
	*l*		*o*
	m		*p*
	*e*		*a*
	*s*		*t*
			r
			a

Figure 3: Champollion's decipherment of Bankes' Ptolemy and Cleopatra cartouches

He then looked at other cartouches to see if he could generate recognisable names from them by applying these sound values. The first one he tried was:

He was able to spell this out as *al?se?tr?*, which seemed to match the Greek *alksentrs* (Alexander), thus giving him three more sign values. Further cartouches both confirmed his values and gave new ones. One nagging thought was that only foreign names might be spelt phonetically but, among others, this cartouche  showed that this was not the case. Champollion knew that the ibis,  , was the symbol of the god Thoth and he read the cartouche as *Thoth-mes*, an old Egyptian name.

As he matched more signs with sound values he was increasingly able to read the hieroglyphic texts as well as the names in the cartouches, and eventually could identify the Egyptian language as Coptic. Champollion's work laid the foundation for Egyptology as it is known today.

3.3 Linear B

The *Linear B* script was a syllabary that was used during the period approximately between 1600 and 1200 BC. Most of the examples come from Crete, particularly Knossos, but there are some from the Greek mainland.

The script consists of some 60 basic signs, 16 optional signs, and about 11 signs that have yet to be deciphered. The script also had signs for numbers (1–1000), and signs for various kinds of weights and measures. There were also sets of signs for different kinds of animals, such as horses and pigs, and for trade goods, such as pots or wool.

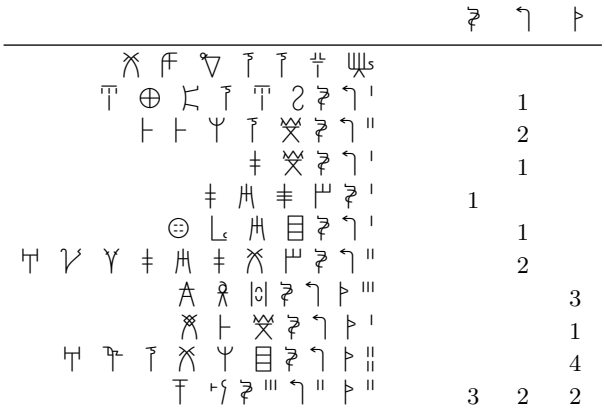


Figure 4: Example of a Linear B text and partial interpretation

Clay tablets bearing the script were found by Sir Arthur Evans (1851–1941) while excavating the ruined Minoan palace at Knossos in Crete, starting in 1900. The tablets were usually small enough to be held in the hand, the largest being about six inches across, ten high and an inch thick. The tablets were accountancy records of some kind and he did work out their numeric systems but not much more than that. As an example, Figure 4 shows, on the left side, the text of a fairly typical tablet. On the right side is an interpretation of , , and , which look rather like an addition sum, where , , and might be units in a non-metric system, like fluid

	Word A	Word B
Case 1		
Case 2		
Case 3		

Figure 5: Two of Alice Kober's triplets

ounces, pints and gallons.[6] With a modicum of effort it can be shown that  = 3 = 18  .

There were many tablets like these where the last line started with either or  , and it was reasonable to assume that these words meant something like 'total'.

On one tablet to do with listings of horses Evans noticed a pair of signs,  , which matched the Cypriot signs  reading *po-lo* (see Table 8), similar to the Greek *polos* for foal. He was convinced that the Cretans spoke an unknown language, which he called Minoan, a theory that he held throughout his life, going to great lengths to disparage anyone who did not agree. Evans guarded his finds somewhat jealously and made little publicly available for others to work on. It was not until 1952, well after his death, that descriptions of his tablets were published.

However, another trove of Linear B tablets had been unearthed at Pylos, on the Greek mainland, by the American Carl W. Blegen in 1939. These were published in 1951 and would probably have been available earlier if the world had not been consumed with the other events of 1939 and later.

Michael Ventris, a British architect, had been fascinated by Linear B since he was a schoolboy and devoted much of his spare time in trying to decipher it. Initially there was no success because of the paucity of material to work on, but the publication of the Pylos tablets changed that.

In the meantime the script had been analysed by various scholars, the signary had been established, and lists had been made of which signs were most common at the start and end of words, and of how signs tended to group themselves. Dr. Alice E. Kober (1907–1950), a classicist at Brooklyn College, had noticed groups of signs where all but the last one or two signs in a word were the same, and thought that this might mean that the language captured by Linear B was inflectional, like Latin or occasionally English as in 'I write' but 'he writes'. In particular she noted words that appeared in three forms, as illustrated in Figure 5, which became known as 'Kober's triplets'.

Ventris had done his own analysis of the script and had come to the conclusion that it was a syllabary. He argued that the difference between the words  or  for total might be due to gender differences in an inflectional language as the first form occurs with the ideogram for man, and the other with the ideogram for woman. If this were the case then the consonants in and were probably the same but the vowels were different. By analysing a number of words in this way he was able to start building up a grid where the signs in each row had the same consonant and those in the same column had the same vowel. It was still a long road, though, from having the signs coordinated in this fashion to being able to read them.

From Kober's work he noticed that there were groups on the Knossos tablets that were not on the Pylos tablets, and made a bold leap to thinking that they might be the names of places on Crete. He suggested that the signs  might be the word for the Greek *Knosos* (Knossos) and  could be the word for the Greek *Amnisos*, the port for Knossos, and a few other names. When he applied the guesstimated values to the signs in the grid he was able to assign values to other signs and start 'reading' a few things. For example, and read *to-so* and *to-sa*, which were similar to the Greek *tosos* (masculine) and *tosa* (feminine) for 'so much' or 'so many'. Ventris had originally whole-heartedly agreed with Evans that the language of Linear B was not Greek, but it was now appearing as though it might well be, especially if the Cypriot *polos* clue was included.

Ventris was familiar with the Greek of Homer (about the ninth century BC) from school but the Linear B tablets were much older than that. The words he was reading seemed similar to Homeric Greek but were not the same. For example there were several tablets from Pylos listing numbers of women, from the ideograph, often followed by two other words,  and  , also with numbers and it was reasonable to assume that these words might be equivalent to 'boys' and 'girls'. However he read them as *ko-wa* and *ko-wo*, whereas the Greek that he knew was *kourai* and *kouroi*. In general the Linear B spellings appeared incomplete, and even when filled out only close to Homeric Greek.

At this point he formed a partnership with John Chadwick, a lecturer in Classics at Cambridge University whose speciality was the early history of the Greek language. Together they worked out a consistent set of rules describing how Greek had changed

[6] The British and the Americans agree that there are eight pints to one gallon, but, perhaps to the chagrin of Texans, there are 20 fluid onces to a British pint and only 16 to an American pint.

Table 4: The basic Linear B syllabary

	a	e	i	o	u
					
d					
j					
k					
m					
n					
p					
q					
r					
s					
t					
w					
z					

Table 5: The Proto-Semitic signary

Name	Glyph	Sound	Meaning
alpu		’	ox
betu		*b*	house
		g	throw stick?
		d	fish?
		z	
		h	man?
wawwu		*w*	hook, peg
hotu		*ḥ*	fence
		ṭ?	twisted flax
yadu		*y*	hand, arm
kappu		*k*	palm of hand
lamdu		*l*	ox goad
mayyuma?		*m*	water
nahasu		*n*	snake
enu		‘	eye
		s?	
		p	leg/foot?
		ṣ?	plant?
		q?	knot?
rasu		*r*	head
		š	lotus pool?
		ʔ	
tawwu		*t*	mark

between Mycenaean and Homeric times. In other words, if you took an arbitrary Linear B tablet, deciphered it and then applied their rules the result would be Homeric Greek. In 1953 they jointly summmarised their work in an article entitled ‘Evidence for Greek dialect in Mycenaean archives’ in *The Journal of Hellenic Studies*. Their theory was completely unexpected and its reception was mixed, to say the least. However it was soon dramatically confirmed.

The American excavation at Pylos had resumed in 1952 after the break for the Second World War. More Linear B tablets were found and stored for later reading. In the spring of 1953 the leader of the team, Carl Blegen, returned to Greece armed with an advance copy of Ventris and Chadwick’s article. Among the newly found tablets was a large one with pictures of three-legged cauldrons, pictures of a number of jars with differing numbers of loops (handles) on top, and Linear B inscriptions. When Blegen applied Ventris’ decipherment he read *tr-ri-po-de*, almost identical to the Greek *tripodes* for three-legged cauldron. Next to the jars with three loops he read *ti-ri-o-we-e* or *ti-ri-jo-we*, and by the jars with four loops *qe-to-ro-we*. The Greek for three in compounds is *tri-* and experts in archaic Greek could accept that *quetro-* would be four in compounds. Once the ‘tripod’ tablet became known most scholars accepted that Ventris had deciphered Linear B, although a few die-hards even went so far as to suggest that the tripod tablet had been ‘planted’ at Pylos!

Table 4 shows the signs in the basic syllabary. Although Linear B was used for writing Greek, there is no other relationship between this ancient script and the Greek alphabet.

3.4 Proto-Semitic

Around 1600 BC there were alphabetic scripts in use in the Middle East that are variously called Proto-Siniatic, Proto-Canaanite, etc. I have lumped these together into a Proto-Semitic font. Several of the signs in this alphabet are obviously derived from Egyptian hieroglyphs, and it may have been a precursor to the Phoenician script.

The alphabet consisted of 23 letters, some of which had alternate forms. Writing was generally from left to right, but could be vertical or in other directions. Table 5 shows the signary, although there is not a complete consensus on this.

4 Phoenician

The Phoenicians initially wrote right to left or left to right. The alphabet consisted of 22 letters although a 23rd glyph was used as the *vav* (or *vau*) character, which had two forms: Y and F. Around 1100 BC the Phoenician alphabet had stabilised and the writing direction was finally fixed as right to left.

Table 6: Evolution of the Phoenician script

Name	Meaning	Hiero.	Proto.	Phoen.	Sound
aleph	ox				’
beth	house				*b*
gimel	camel				*g*
daleth	door				*d*
he	window?				*h*
vav	nail			F, Y	*w*
zayin	dagger?		=	I	*z*
heth	fence?				*ḥ*
teth				⊗	*ṭ*
yod	hand			I	*y*
kaph	palm of the hand				*k*
lamed	ox goad				*l*
mem	water				*m*
nun	fish				*n*
samekh	prop or post				*s*
ayin	eye			○	‘
pe	mouth	?			*p*
sade					*ṣ*
qoph	knot?				*q*
resh	head				*r*
shin	teeth				*š*
tav	mark or cross		+	+	*t*

Table 6, which is somewhat speculative in some cases, illustrates how the Phoenician script (may have) developed from the Proto-Semitic script; the name and, where known, the meaning of the Phoenician glyph is given, as is the transliterated sound value. It also shows how some of the Proto-Semitic glyphs may have been inspired by the Egyptian hieroglyphs; in some cases the derivation is obvious.

5 Later Western scripts

Among the later Western scripts — those developed after 1000 BC — Greek, Etruscan and Latin are the direct ancestors of the English alphabet. Others, such as Cypriot and Runic, are off by themselves.

5.1 Greek

Initially the Greeks used the Phoenician alphabet and also wrote right to left but by the 7th century BC it became boustrophedon and around 500 BC they finally settled on writing left to right. The Greeks added new letters to the Phoenician abecedary so that around the 6th century BC their alphabet consisted of 26 characters. The Y form of the Phoenician *vav* became the Greek *upsilon* while the F form of *vav* became the Greek *digamma*. The names of the letters lost their meanings and instead effectively stood for the pronunciation of the letter. The Greeks also added the *psi*, *phi* and *omega* characters. Several different glyphs were used for each character, depending on geographical location, whether on the mainland or around the Aegean Sea. One variety of the 6th century BC alphabet looked like this:

ΑΒΓΔΕFΙΗΘΙΚΛΜΝΞΟΓΡΡΣΤΥΧΦΥΩ

In 403 BC the Athenian citizens codified the alphabet with the glyphs looking much as they do today. The *digamma* and the *qoph* characters were dropped from the abecedary, thus leaving the 24 characters that we are now accustomed to. The 4th century BC alphabet was like this:

ΑΒΓΔΕΖΗΘΙΚΛΜΝΞΟΠΡΣΤΥΧΦΨΩ

5.2 Etruscan

The Etruscans, forerunners of the Romans in Italy, based their alphabet on the Greek abecedary, but they continued to write right to left as the Phoenicians had, so their glyphs were mirrored with respect

Table 7: The changing alphabets

Phoenician	Greek			Etruscan	Roman	Modern
	600 BC	394 BC	Modern			
𐤀	Α	Α	Α	Α	Α	A
𐤁	Β	Β	Β	Β	Β	B
𐤂	Γ	Γ	Γ	Ⅽ	C G	C G
𐤃	Δ	Δ	Δ	D	D	D
𐤄	Ε	Ε	Ε	Ε	Ɛ	E
𐤅	Υ	Υ	Υ		V Y	U V W Y
𐤆	Ϝ			Ϝ	Ϝ	F
𐤇	Ι	Ζ	Ζ	Ι	Ζ	Z
𐤈	Ħ	Η	Η	Ħ	Η	H
𐤉	⊗	⊙	Θ	⊗		
𐤊	Ι	Ι	Ι	Ι	Ι	I J
𐤋	Κ	Κ	Κ	Κ	Κ	K
𐤌	Λ	Λ	Λ	Ⅼ	Ⅼ	L
𐤍	Μ	Μ	Μ	Μ	Μ	M
𐤎	Ν	Ν	Ν	Ν	Ν	N
𐤏	Ξ	Ξ	Ξ	⊟	Χ	X
𐤐	Ο	Ο	Ο	Ο	Ο	O
𐤑	Γ	Γ	Π	Γ	Ρ	P
𐤒				Μ		
𐤓	Ϙ			Ϙ	Ⴕ	Q
𐤔	Ρ	Ρ	Ρ	Ρ	Ꭱ	R
𐤕	Σ	Σ	Σ	Ϟ	Ꮪ	S
𐤖	Τ	Τ	Τ	Τ	Τ	T
	Ⓘ	Φ	Φ	Φ		
	Ξ	Ξ	Χ			
	Ψ	Ψ	Ψ	Ψ		
	Ω	Ω	Ω			

to the Greek ones. The Etruscan script was used until the first century AD, although the Etruscans had disappeared well before then.

The Etruscan abecedary originally consisted of 26 characters but by about 450 BC it had decreased to only 20. One interesting glyph that was dropped looked like the digit 8, and denoted an 'f' sound. Some scholars have surmised that this was the ancestor of the mathematical sign for infinity (∞). In the 8th century BC the alphabet looked like:

Α Β Γ Δ Ε Ϝ Ι ⊟ ⊗ Ι Κ Ⅼ Μ Ν Ο Γ Μ Φ Ρ Ϟ Τ Υ Χ Φ Ψ 8

As far as decipherment is concerned, Etruscan is a known script used for writing an unknown language. Scholars are able to read aloud the script but they do not understand the language. Apart from proper names only about a couple of dozen words are known, and these are mainly to do with family relationships like 'son' and 'father'.

5.3 Latin

The Romans based their script on the Etruscan one, again adding and dropping characters. They added the G and Y characters but dropped *theta*, *psi*, and *phi* to end up with a 23 character abecedary, although they rarely used the H, K, and Z characters. So, their alphabet looked like:

A B C D E F G H I K L M N O P Q R S T V X Y Z

The lettering used in this last example is a copy of the capitals engraved on the Trajan Column in Rome which was erected in 114 AD. Many typographers believe that these represent the high point in the Roman artistic legacy. Unfortunately they do not reproduce well in the size shown here. In real life the inscription is about six feet above eye level and there are six lines of text. The letters on the top line are somewhat over four inches tall, decreasing to about three inches on the lowest line, which presents the illusion that they are all of the same height when observed from the normal viewpoint.

Summarising, Table 7 shows the 1100 years of development of the Latin alphabet.

Table 8: The Cypriot syllabary

	a	e	i	o	u
					
g					
j					
k					
l					
m					
n					
p					
r					
s					
t					
w					
x					
z					

5.4 Cypriot

The Cypriot script was a syllabary used in Cyprus during the approximate period between 1000 and 200 BC for writing Greek. It has a relationship to Linear B as it includes some of the same signs. Towards the end of its life few people could read the script, so inscriptions were written using both the syllabary and the Greek alphabetic characters. These bilinguals made it relatively easy to decipher the script, a task that was essentially completed by the last quarter of the nineteenth century.

Like Linear B, the Cypriot script has no relationship with the Greek abecedary apart from the fact that both can be used for writing the same language.

Table 8 shows the Cypriot syllabary.

5.5 Runic

The runic alphabet, which is not shown in Table 7, is known as *futhark* after its initial characters. It was used, with local variations, in the Germanic, Scandinavian and Anglo-Saxon countries until shortly after printing was invented. Scholars are unclear as to the origins of the futhark abecedary, but there are obvious correspondences between some of the glyphs and the Phoenician and Etruscan ones, while others have no resemblance at all.

Like the Phoenician alphabet, the names of the futhark characters have meanings. The ordering of the characters, together with their names and meanings, is shown in Table 9.

It is very noticeable that the letter ordering is completely different from any of the other abecedaries in Table 7. It is interesting to speculate whether the ordering of an original abecedary depends on the frequency of use of the letters, or those with the most important meanings have priority.

The *wen* character (ƿ) is no longer used in English, but does indicate that the Anglo-Saxons had need of a 'W'. The *thorn* character (þ) is like *theta* in that it represents the 'th' sound. Early printers usually did not have a þ, so they used a 'Y' character instead. From this practice comes the modern affectation of naming something like 'Ye Olde Pub' instead of 'The Old Pub'. Also, it has the *ger* character (ȝ) which corresponds to the modern 'J' sound — 'J' did not appear in the Latin alphabet until about the mid-1500's.

6 Later Semitic scripts

This section includes scripts that were invented after 1000 BC and used in the Middle East.

Table 10 shows the evolution of the modern Hebrew and Arabic[7] scripts.

6.1 Old Persian

It is believed that the Old Persian cuneiform script was invented on the order of the Persian king Darius I for use on royal monuments. The script was only in use between about 500 and 350 BC.

Old Persian was a syllabary with 36 glyphs. There were also 5 ideographs, some with multiple forms, for the words *king*, *country*, *earth*, *god* and *Ahuramazda* (the Persian god), together with a word divider. Numerals were also represented.

Somewhat surprisingly the decipherment of Old Persian led directly to the decipherment of the far older Sumerian and Babylonian cuneiform scripts. The basic work was done by Georg Friedrich Grotefend (1775–1853), a high school teacher in Göttingen and was completed by Henry Rawlinson [Adk04]. It was generally assumed that because of the limited number of signs the script was alphabetic, the slanting wedge was probably a word divider and it was written left to right.

Grotefend started with two texts which were inscribed above doorways in the ruined city of Persepolis. The first was:

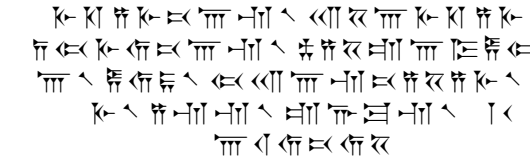
[7] Arabic actually has 28 characters and many more glyphs as they change depending on the position of the character in a word, but I have only shown the characters that were derived from the Phoenician script.

Table 9: The Futhark abecedary

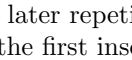
Glyph	Name	Meaning	Glyph	Name	Meaning
ᚠ	*feof, feh, fe*	wealth	ᛇ	*hic, ih, eoh*	
ᚢ	*ur, hur*	auroch	ᛈ	*peord*	
ᚦ	*thorn*		ᛉ	*eohlx*	
ᚫ	*æsc, os*	oak tree	ᛋ	*sigel*	sun
ᚱ	*rad, rat*	riding	ᛏ	*tir*	name of a star?
ᚳ	*ce, kaun*	torch	ᛒ	*berc, birth*	birch tree
ᚷ	*gebu, gyfu*	gift	ᛖ	*hæc, ech, eh*	horse
ᚹ	*wen*	joy	ᛗ	*man*	man
ᚻ	*hegl, hagal*	hail	ᛚ	*lagu*	water or sea
ᚾ	*nyd, nod*	need or hardship	ᛝ	*ng*	
ᛁ	*is*	ice	ᛞ	*dag, dæg*	day
ᛄ	*ger, yr, ar*	year	ᛟ	*o, oe*	mouth
			᛫		punctuation

Table 10: Evolution of Middle Eastern scripts

Name	Hiero.	Proto.	Phoen.	Aram.	Nab.	Hebrew	Arabic
aleph	𓄿	𐤀	𐤀	𐡀	𐢀	א	ا
beth	𓃾	𐤁	𐤁	𐡁	𐢁	ב	ب
gimel	𓌙	𐤂	𐤂	𐡂	𐢂	ג	ج
daleth	𓂝	𐤃	𐤃	𐡃	𐢃	ד	د
he	𓀠	𐤄	𐤄	𐡄	𐢄	ה	ه
vav	𓇼	𐤅	𐤅, 𐤉	𐡅	𐢅	ו	و
zayin		𐤆	𐤆	𐡆	𐢆	ז	ز
heth	𓉔	𐤇	𐤇	𐡇	𐢇	ח	خ
teth	𓍑	𐤈	𐤈	𐡈	𐢈	ט	ط
yod	𓂧	𐤉	𐤉	𐡉	𐢉	י	ى
kaph	𓎡	𐤊	𐤊	𐡊	𐢊	כ	ك
lamed	𓌹	𐤋	𐤋	𐡋	𐢋	ל	ل
mem	𓈖	𐤌	𐤌	𐡌	𐢌	מ	م
nun	𓆓	𐤍	𐤍	𐡍	𐢍	נ	ن
samekh		𐤎	𐤎	𐡎	𐢎	ס	س
ayin	𓁹	𐤏	𐤏	𐡏	𐢏	ע	ع
pe	𓂋 ?	𐤐	𐤐	𐡐	𐢐	פ	پ
sade		𐤑	𐤑	𐡑	𐢑	צ	ص
qoph		𐤒	𐤒	𐡒	𐢒	ק	ق
resh	𓁶	𐤓	𐤓	𐡓	𐢓	ר	ر
shin	𓈙	𐤔	𐤔	𐡔	𐢔	ש	ش
tav	𓏏	𐤕	𐤕	𐡕	𐢕	ת	ت

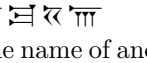


And the second was:

At the time of the Achaemenid Dynasty (559–331 BC) in Persia the Zoroastrians were prominent together with their sacred text, the *Zend-Avesta*. This had recently become known in Europe as well as some Pehlevi[8] texts translated and published by Silvestre de Sacy in 1793 in his *Mémoires sur diverses Antiquités de Perse*. Grotefend knew of a formula from these texts that went 'X, king . . . , king of kings, son of Y. . . '. He guessed that the Persepolis inscriptions might embody the same formula and that the second sign group in each inscription () might be the word for king, and the later repetitions equate to 'king of kings'.

In the first inscription the first group

should be the name of a king, say A for the sake of argument, and similarly in the second inscription the first group



would be the name of another king, say B. Grotefend noted that A also occurred in B's inscription, so that the second one might read 'B, king. . . , king of kings, son of king A. . . '. He guessed again that the inscriptions might have something to do with Darius and Xerxes, and in which case A would be Darius and B his son Xerxes.

Using forms of Darius and Xerxes which he derived from an amalgamation of Avestan, Greek and Hebrew he suggested that the signs in the two names should be read as *d/a/r/h/e/u/sh* for Darius and *kh/sh/h/e/r/sh/e* for Xerxes. With these values the signs for king would read *kh/sh/e/h/?/?/h*. In the Avesta he found the kingly title *khscheio* and he took this as confirmation that he was on the right track and the language of the inscriptions was Avestan. He was correct.

With the limited number of inscriptions available, work could not proceed further. However, huge inscriptions were discovered on the side of a mountain near Behistun in western Iran. Henry Rawlinson (1810–1895), who was serving as the British military advisor to the brother of the Shah of Iran, began to copy the inscriptions in 1835. Among other things, this involved dangling from ropes to get at some of the texts as they were on a cliff more than 300 feet up. In all, it took him ten years to copy 414 lines. It turned out that the inscriptions were trilinguals, in Old Persian, Elamite and Babylonian. Rawlinson effectively completed the decipherment of Old Persian which then enabled him to tackle the Babylonian, which became the key to other cuneiform scripts.

[8] Pehlevi (or Pahlevi or Pahlavi) was a Persian dialect of the Sassinide period (3rd–7th century AD).

Table 11: The Old Persian syllabary

	a	i	u		a	i	u
k				f			
x				b			
g				m			
c				y			
j				r			
t				l			
th				v			
ça				s			
d				ša			
n				z			
p				h			

Some of the other glyphs include: (king), (country), (earth), (god), and (Ahuramazda).

Examples of numerals include: (1), (2), (10), (20), (100).

As a final note, the transliteration of the second of the two inscriptions that Grotefend tackled is:

xa-ša-ya-a-ra-ša-a-: xa-ša-a-ya-tha-i-ya-: va-za-ra-ka-: xa-ša-a-ya-tha-i-ya-: xa-ša-a-ya-tha-i-ya-a-na-a-ma-: da-a-ra-ya-va-ha-u-ša-: xa-ša-a-ya-tha-i-ya-ha-ya-a-: pa-u-ça-: ha-xa-a-ma-na-i-ša-i-ya-:

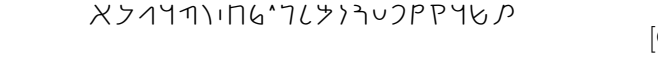
and which, when translated, means:

Xerxes, the great king, the king of kings, the son of Darius the king, an Achaemenian.

6.2 Aramaic

The Aramaic script is an early offshoot from the Phoenician and was used between about the tenth and second centuries BC in the Middle East. The Aramaic script also branched and led to both modern Arabic and Square Hebrew scripts.

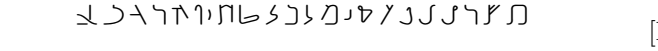
The script is alphabetical and consists of 22 consonants.



6.3 Nabatean

The Nabatean script is an offshoot of the Aramaic script and was in use in an area centered around Petra — the 'rose-red city half as old as time' — roughly during the period between the fourth century BC and the fourth century AD. It is a direct ancestor of the modern Arabic script.

Like other Semitic scripts it is alphabetical and consists of 22 consonants.



7 Remarks

The result from formula 1 is certainly an approximation. I applied it to the two Old Persian texts in section 6.1 which together contain 150 signs with 22 different kinds. The estimated number of signs is

$$S = 150^2/(150 - 22) - 150 = 25.78$$

which is somewhat under the actual value of 36 for the syllabary. The much shorter made up text on page 202 consisting of 70 characters of 29 different kinds gives

$$S = 70^2/(70 - 29) - 70 = 49.5$$

which is a significant overestimate. However, combining the three texts gives 220 total characters with 33 different kinds, resulting in

$$S = 220^2/(220 - 33) - 220 = 38.8$$

which is close to the actual number.

The books listed below are among the more accessible sources describing the development of the alphabet and the Latin script, and of decipherments of archaic scripts.

The fonts used in this article can be obtained from CTAN (the Comprehensive TEX Archive Network). The Arabic script came from Klaus Lagally's `arabtex` package in the `languages` area, and similarly the Hebrew script is from `hebrew/hebtex` in the same area. The Trajan font is in the `fonts/trajan` directory. All the other scripts are in the `fonts/archaic` directory.

References

[Adk04] Lesley Adkins. *Empires of the Plain: Henry Rawlinson and the Lost Languages of Babylon*. Harper Perennial, 2004.

[Bud89] E. A. Wallis Budge. *The Rosetta Stone*. Dover, 1989.

[Cha58] John Chadwick. *The Decipherment of Linear B*. Cambridge University Press, 1958.

[Cha87] John Chadwick. *Linear B and Related Scripts*. Reading the Past. University of California Press, 1987.

[CM98] Mark Collier and Bill Manley. *How To Read Egyptian Hieroglyphs*. University of California Press, 1998.

[Coe99] Michael D. Coe. *Breaking the Maya Code*. Thames & Hudson, 1999.

[Dav87] W. F. Davies. *Egyptian Hieroglyphs*. Reading the Past. University of California Press, 1987.

[Fri89] Johannes Friedrich. *Extinct Languages*. Dorset Press, 1989. (Translated from the original German *Entzifferung Verschollener Schriften und Sprachen* by Frank Gaynor).

[Gor87] Cyrus H. Gordon. *Forgotten Scripts: Their Ongoing Discovery and Decipherment*. Dorset Press, 1987.

[Hea90] John F. Healey. *The Early Alphabet*. Reading the Past. University of California Press, 1990.

[Ifr00] Georges Ifrah. *The Universal History of Numbers*. John Wiley & Sons, 2000. (Originally published as *Histoire universelle des chiffres*. Robert Laffort, Paris, 1994.).

[Mal91] J. P. Mallory. *In Search of the Indo-Europeans: Language, Archaeology and Myth*. Thames and Hudson, 1991.

[Nak80] Akira Nakanashi. *Writing Systems of the World: Alphabets, Syllabaries, Pictograms*. Charles E. Tuttle Company, 1980.

[Pop99] Maurice Pope. *The Story of Decipherment: From Egyptian Hieroglyphs to Maya Script*. Thames & Hudson, 1999.

[Rob95] Andrew Robinson. *The Story of Writing: Alphabets, Hieroglyphs & Pictograms*. Thames & Hudson, 1995.

[Rob02] Andrew Robinson. *Lost Languages: The Enigma of The World's Undeciphered Scripts*. McGraw-Hill, 2002.

[Wal87] C. B. F. Walker. *Cuneiform*. Reading the Past. University of California Press, 1987.

◇ Peter Wilson
18912 8th ve. SW
Normandy Park, WA 98166
USA
herries.press (at) earthlink.net