

Written and spoken language



Linguists generally agree with the following statement by one of the founders of the modern science of language.

"Writing is not language, but merely a way of recording language by visible marks."

Leonard Bloomfield, *Language* (1933)

Another way to express Bloomfield's point is to say that writing is **"parasitic" on speech**, expressing some but not all of the things that speech expresses. Specifically, writing systems convey the sequence of known words or other elements of a language in a real or hypothetical utterance, and indicate (usually somewhat less well) the pronunciation of words not already known to the reader. Aspects of speech that writing **leaves out** can include emphasis, intonation, tone of voice, accent or dialect, and individual characteristics.

Some caveats: Writing is usually **not** used for "recording language" in the sense of transcribing speech. Writing may substitute for speech, as in a letter, or may deploy the expressive resources of spoken language in visual structures (such as tables) that can't easily be replicated in spoken form at all. In addition, writing systems may include some **conventions** that are substantially independent of speech. Most basically, words that are pronounced the same way may not be written the same way ("homophones"), and vice versa ("homographs"). More generally, the written form of a language may represent a dialect with special vocabulary and grammatical principles; but this develops only secondarily, once writing has been established in a spoken context.

Still, Bloomfield was basically correct: writing is a way of using "visible marks" to **point** to pieces of real or potential spoken language.

Primacy of spoken language



Some version of this is clearly true, as we can see by looking at the **history of the human species and of each human individual**. In both regards, **spoken language precedes written language**.

Speech

Present in every society

Human evolution has led to adaptations for speaking

Learned before writing

Learned by all children in normal circumstances, without instruction

Writing

Present only in some societies, and only rather recently

Evolution has not specifically favored writing

Learned after speech is acquired

Learned only by instruction, and often not learned at all

Don't confuse sound and spelling



One of the points of studying phonetic transcription, as we did last week, is to separate sound and spelling. This is a very common confusion among non-specialists.

A *New York Times Magazine* article (November 6, 2005), about the **use** of the word *filadelfia* in Spanish to refer to cream cheese, mentions some Kraft commercials in the 1990's that led to this association:

One of the spots played off the lack of the "ph" sound in Spanish. It featured a little girl who inquisitively sounded out the name on the package: "*Pee-la-del-pia?*" she asked.



The author is confused about sound versus spelling. Spanish has the sound [f], which is what "ph" represents; but it lacks the orthographic principle mapping from "ph" to [f]. Conveniently, since "h" is silent in Spanish, the child easily ignores it and pronounces [p] instead.

Types of writing



Here are **traditional terms** for different types of writing systems.

- Ideographic ("idea-writing")
- Logographic ("word-writing")
- Syllabic
- Alphabetic

In fact they're best as ways to talk about particular **symbols** rather than entire systems, since particular systems often mix types of symbols. We'll consider each one in turn.

Ideographic symbols



An ideographic sign refers to an **idea** or general concept, not a specific word of some language; therefore this is **not** true (linguistic) writing. In the absence of invented or borrowed writing, societies use ideographic symbols exclusively.






The **Dakota Indians** of the northern plains of the U.S. on occasion used a series of symbols, one for the passage of each year (or "winter"), to keep track of events beyond the memory of living individuals. A famous example of such a **winter count** was maintained by Lone Dog of the Yankton band of Dakotas, living in Montana; it runs from 1800-1871. The symbols begin in the center of a buffalo hide and spiral outward.

Symbols in the Winter Count



Below are a few examples of symbols used in Lone Dog's Winter Count. These symbols are examples of ideographic **proto-writing**, since they are not connected to particular words, but rather to events or ideas that are suggested by the image.

<p>1856-57</p>		<p><i>Four-Horn is made a medicine-man</i> (man with four horns holds ceremonial pipe)</p>
<p>1861-62</p>		<p><i>Buffalo so plentiful that they come close to the dwellings</i> (tracks near tepee)</p>

1869-70		<i>Solar eclipse</i> (black sun, red stars)
---------	--	--

Symbols that are pictures (more or less) of the thing they represent are called **pictograms**. Most ideograms are also pictograms. Each design serves to remind the interpreter of the count of an event, but much of the detail has to be provided by memory. In other words, it is a **mnemonic**.

Garrick Mallery, *Picture-Writing of the American Indians*, 1893, pp. 266-287.

Ideographic symbols in modern life ← ↑ →

In societies with writing, ideographs are still quite useful, particularly when a **symbol independent of a particular language** is desired, or where it will be quicker or easier to grasp a graphic symbol than written words. Road signs and other public signs are increasingly ideographic in today's world.

The written word **restroom** would be of use only to a (literate) speaker of English, whereas this symbol is easily learned regardless of a person's linguistic background. The kinds of messages that can be conveyed without reference to language are quite limited, however.

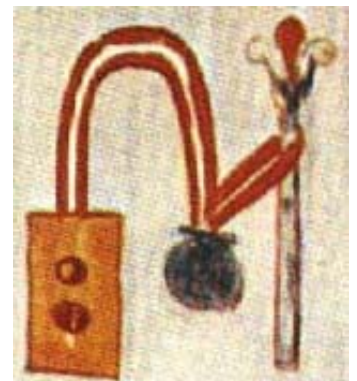
Ideographic systems as such have never developed into a form fully capable of conveying unlimited messages from one person to another. Instead, they either remain as **limited systems operating within a highly restricted application** — say, to keep warehouse records — or else they develop into a genuine writing system, capable of conveying any linguistic message. In the second case, the process of development into a genuine writing system always involves adding some phonetic aspects, in ways we'll describe shortly.



An Egyptian logogram: *ziçiR* "write" ← ↑ →

A logographic sign refers to a **specific word**, not just the general idea behind that word. The notion of "word" is not entirely correct here; instead, the right idea is **morpheme**, or the minimal unit of meaning. We'll learn more about this when we talk about morphology in two weeks. In simple cases, however, words and morphemes are the same thing. All the **early** writing systems of antiquity are heavily logographic, because what happens at the beginning of writing is a shift from idea to word as the element being represented by a symbol.

For example, this depiction of a scribe's tools stands for the **Egyptian** word originally pronounced something like **ziçiR**, meaning "write." (Some details of the earliest Egyptian pronunciation are subject to debate.) It is a **logogram** (or "logograph").



A Chinese logogram: *mǎ* "horse"



Similarly, this Chinese symbol (when read in Mandarin) stands for the word *mǎ* (with a falling-rising tone on the syllable), meaning "horse."

马

It's a simplification of what began as a drawing of a horse: here are some historical stages in the form of the sign.



More Chinese logograms



Nothing about these representations indicates the **pronunciation**; this lack is one of the basic characteristics of a logogram as an individual symbol. Some more Chinese examples:

人	木	火	水	上	下
<i>rén</i>	<i>mù</i>	<i>huǒ</i>	<i>shuǐ</i>	<i>shàng</i>	<i>xià</i>
"person"	"tree"	"fire"	"water"	"above"	"below"

While individual **symbols** can be logographic, all full writing **systems** involve reference to pronunciation in some way — including Chinese for the great majority of its characters, as we'll see below.

Syllabic writing in Japanese



A syllabic sign refers to a **syllable**, typically a consonant plus a vowel.

An example of a script in which each symbol normally represents a syllable is the **hiragana** syllabary of Japanese. Here's the word *watakushi*, meaning "I" (a polite form).

わ た く し
wa ta ku shi

Hiragana



Notice that each symbol stands for a consonant plus a vowel — just four symbols are needed, as opposed to the nine used to write it in English letters. The basic set of syllabic symbols is shown in this chart.

	k	s	t	n	h	m	y	r	w	N
a	あ	か	さ	た	な	は	ま	や	ら	わん
i	い	き	し	ち	に	ひ	み	り		
u	う	く	す	つ	ぬ	ふ	む	ゆる		
e	え	け	せ	て	ね	へ	め	れ		
o	お	こ	そ	と	の	ほ	も	よろ	を	

The voiced obstruents /g, z, d, b/ are indicated by adding two small strokes to the symbol for the corresponding voiceless consonant; and for historical reasons the /p/ symbols are derived from the /h/ symbols by adding a small circle (historically, it was actually [p] that became [h]). The N symbol is used when /n/ occurs in the coda of a syllable. A parallel syllabary called *katakana* is used for borrowed vocabulary. Japanese also makes pervasive use of Chinese characters, in ways too complex to describe here.

Morphosyllabic components of *mā* "mother"



While some basic symbols in Chinese are truly logographic, the great majority of characters actually consist of **two parts**: one that tells about **meaning**, and one that indicates (with varying accuracy) the **pronunciation**. So the *mǎ* "horse" character above functions as a phonetic indicator in the following character for "mother," which points to the spoken word *mā* with a high tone. The left side of this character means FEMALE.

The character:

妈

Its parts:

FEMALE 妈 [ma]

Since the "horse" component indicates the pronunciation of the whole syllable [ma], it is a **syllabic** symbol: there's no subpart that means [m] or [a], for example.

It is clearly inappropriate to call the Chinese system "ideographic", as is sometimes done. Chinese characters refer to **specific morphemes**, not vague ideas. However, Chinese characters are simultaneously a kind of **syllabic** writing, since the phonetic element always is a syllable in size. The term **morphosyllabic** has been created to describe it.

The important point is that Chinese, though in some ways logographic, also has a very large **phonetic aspect** to its symbols — more exactly, a **syllabic** one. The imprecision of the phonetic aspect is analogous to the irregularities of English spelling. That is, English spelling usually tells us what the morphemes are, but unless we know in advance, it often gives us only **imperfect information** about pronunciation. We can be sure that *tough* will not be pronounced *congressional* or *book*, but only knowledge of the word itself tells us that it rhymes with *rough* and not with *dough* or *through* or *plough*.

No one has ever developed a full communications system based on ideographic principles, although people have often surmised that this would be useful, because it would (or at least could) be universal. The problem is that "universality" means only that it is equally hard for everyone to develop and learn such a system. Imagine trying to produce the

precise word choice and sentence structure of a legal contract or a poem if all you can write is "ideas".

The same character as a phonetic and as a semantic



For each composite character, one part is the **phonetic**, and the other is the **semantic**; generally the rest of what that character expresses independently is ignored. Thus in "mother" above, the left character means FEMALE, while the same character, as part of "servant" below, is used for its sound [nü].

Basic character:

女

nǚ

"female"

As a semantic:

妈

mā

"mother"
(with [mǎ])

As a phonetic:

奴

nú

"servant"
(with HAND)

Complex components



The resulting composite character *nú* has a particular pronunciation, of course, so it can also be **used as the phonetic** in other, denser characters. Due to slow historical accretion of new characters, this situation is quite common.

Composite character:

奴

nú

"servant"

Two examples as a phonetic in other characters:

努

nǚ

"work"
(with STRENGTH)

怒

nù

"anger"
(with HEART)

In the majority of cases, seen also here, the phonetic component of the character is an imperfect reflection of the pronunciation (sometimes the tone is different, sometimes the consonant or vowel, sometimes all of them).

Alphabetic writing



An alphabetic sign refers to a single **sound**, whether a consonant or a vowel. This type of script hardly needs illustration, since English uses the Latin alphabet. But there are several aspects of English spelling that depart from a "pure" alphabet. These include **digraphs**, or the use of two letters to write a single sound, as well as rarer trigraphs which use three letters for one sound; and **silent** letters that aren't pronounced at all.

Digraphs (and trigraphs)

Silent letters

- fish, chin, thin
- see, too, saw
- watch

- caught
- give

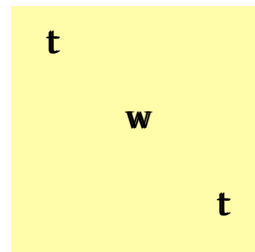
These "impurities" don't change the fact that the **basic nature** of the system is one symbol per sound — that's the standard from which these examples deviate (rather than, say, a syllabic standard).

Egyptian alphabet



There are many other kinds of alphabets as well. One sub-part of the **Egyptian** writing system was alphabetic, though originally it indicated only **consonants** (like the Semitic alphabets that are probably descended from Egyptian).

One example is the word **tw** "image", which we can render as [tut]. The little loaf of bread stands for the sound [t], and the quail chick for the sound [w].




This is the first part of the name *Tutankhamun*, literally "living image of Amun."

Hebrew consonantal alphabet



The modern **Semitic alphabets** work in a similar way. In Hebrew, for example, only certain vowels are written (as secondary functions of consonant characters), and there can be multiple ways of reading a particular spelling. (These alphabets are written from right to left.)

CONSONANTAL SPELLING	ANCIENT	MODERN	MEANING
 r - p - s ←	<i>sāfar</i>	<i>safar</i>	"he counted"
	<i>sēfer</i>	<i>séfer</i>	"book"
	<i>s^efār</i>	<i>sfar</i>	"frontier"
	<i>sappār</i>	<i>sapar</i>	"barber"
	<i>sippēr</i>	<i>siper</i>	"he told"
	<i>suppar</i>	<i>supar</i>	"it was told"

The multiple reading of some letters, such as **פ** *pey* which can be [p] or [f], is due to certain historical changes in the language, leading to a fairly complicated modern situation. Briefly, the stop consonants [p, b, t, d, k, g] were pronounced as fricatives when preceded by a vowel, but not when geminated (doubled). In modern pronunciation, only [p, b, k] still have variants, which are [f, v, x]. Since consonant gemination is no longer observed, we now find minimal pairs such as *safar* and *sapar* which would have been impossible in the ancient language.

Hebrew vowels



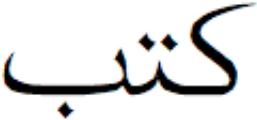
Especially in later forms of writing, including Modern Hebrew, it's typical to use some consonant letters to indicate **vowels**. This can make the reading more explicit, as for the last two *s-p-r* words written with the *y* and *w* consonants for the vowels [i] and [u].

 <p>r - p - y - s ←</p>	 <p>r - p - w - s ←</p>
[siper]	[supar]

Arabic consonantal alphabet



In Arabic, it can be trickier for a novice to identify the letters, since most of them have to be connected up in cursive fashion. Here's one example with three letters, "k-t-b". Since in everyday writing short vowels are not written, and doubled consonants are not distinguished, this word can be read in several ways.

CONSONANTAL SPELLING	PRONUNCIATION	MEANING
 <p>b - t - k ←</p>	<i>kataba</i>	"he wrote"
	<i>kutiba</i>	"it was written"
	<i>kattaba</i>	"he caused to write"
	<i>kutub</i>	"books"

Context generally makes things clear, just as it does for ambiguous English written words like *read* and *bow*. In addition, Arabic consistently writes long vowels with consonant letters, similar to the vowels in the Hebrew examples *siper* and *supar*; so *kitāb* "book" would have an extra letter, *alif*, for the long [ā].

Early writing



The following map shows a range of early writing systems.



Writing was invented perhaps **four different times** in the history of the world, though the exact number is a matter of debate. For example, some believe that Egyptian writing was inspired by contact with Sumer, though the system itself was a new creation; others argue that Egyptian was independent, and perhaps earlier.

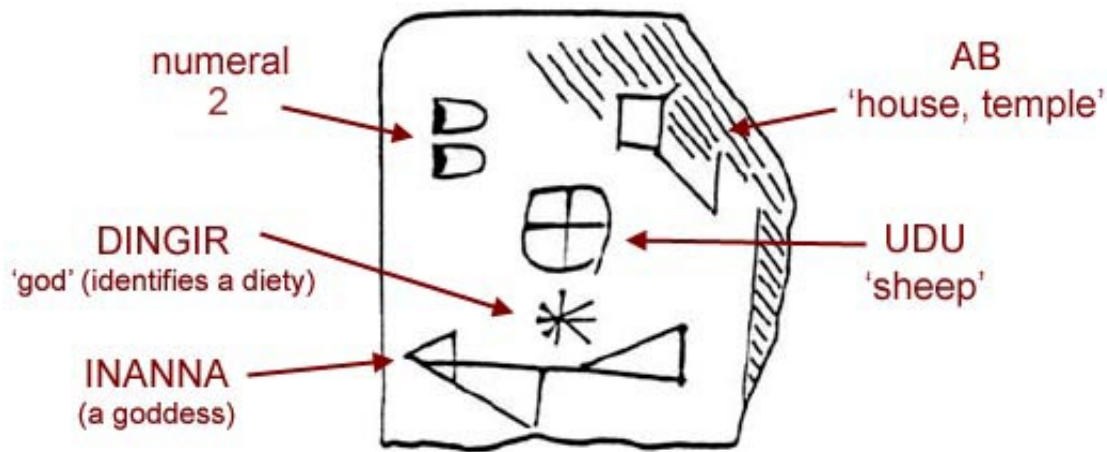
Sumerian	3200 BCE
Egyptian	3250-3050 BCE
Chinese	1200 BCE
Maya	292 CE

The **earliest writing** may have arisen in southern Mesopotamia (modern Iraq), among the Sumerian people, between 3200 and 3000 BCE. Recent discoveries suggest that Egyptian may have been just as early, however.

Early Sumerian tablet



They developed a system of icons inscribed on clay tablets for keeping **temple economic records**. An early example includes icons for "two", "sheep", "house/temple", and the goddess "Inanna" with the general symbol for "god" used to identify divine beings (but not actually pronounced in such situations).



Such early documents are **difficult to interpret**, largely because the symbols seem to be arranged fairly randomly on the tablet. They lack indication of verbs and the interrelations of nouns — precisely what full language excels at. The meaning might be:


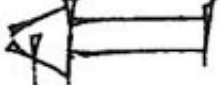



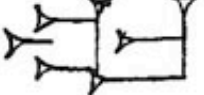





*"two sheep have been received from the temple of Inanna", or
 "two sheep were delivered to the temple of Inanna",
 or perhaps something else entirely.*

Thus these marks constituted a **limited notation system**, which in the beginning may only have served to remind the writer of what he had once already known (a **mnemonic**, like the Dakota winter count). However, as long as agreed-on standards were obeyed, another person could also read the record in the same way.

"Charades"



In this regard, these early documents were similar to many systems for **record-keeping**, based on symbolic tokens of many sorts, developed over and over again in many cultures over the millennia — marks on stone or bone, clay figurines, even knots in cords. As civilizations become more complex, however, record-keeping needs become **increasingly complex** as well. The ability of trained third parties to read such records in a consistent way becomes important for mediating or adjudicating disputes in non-violent ways. In the case of the Sumerian record-keeping system, two **crucial innovations** led (over a few hundred years) to a full writing system.

	Uruk ca. 3000	Lagash ca. 2400	Assyrian ca. 700
SAG 'head'			
KA 'mouth'			
GU 'to eat'			
EME 'tongue'			

A basic pictograph, such as for "head" here, can be used as the basis for writing related words by adding new marks, as for "eat". This is often much easier than creating new signs.

- SAG "head" begins as a **drawing** of a head, though it becomes more abstract as time goes on.
- KA "mouth" is created by making **arbitrary marks** at the location of the mouth, relative to the head. This is like pointing at a subpart of a picture.
- GU "to eat" is created from "mouth" by juxtaposing a bowl from which one would eat. The two meanings "mouth" and "bowl" together suggest the notion of eating. This is a **semantic+semantic** sign, since there's still no reference to pronunciation.
- EME "tongue" is created by adding to the sign "mouth" another sign which has the pronunciation ME. (It means "oracle, function" when used independently, but its meaning is not invoked here.) This composite sign is **semantic+phonetic** because [me] evokes the word EME.

A beginning Sumerian reader was in effect being asked to play a sort of game of **charades**: what word has something to do with "mouth" and sounds like [me]? — why of course, that's EME, "tongue"! This same technique was used for several other signs based on KA "mouth".

- NUNDUM "lip" is composed of "mouth" plus [nun].
- SU "beard" is composed of "mouth" plus [sa] — an imperfect match, but still enough to get the idea across initially.

These combinations quickly became **conventionalized**, however, so that a competent reader of the language did not need to rely on guessing: he simply **memorized** the fact that, for example, EME consists of the same graphic pieces as KA and ME.

Chinese semantic+semantic characters



A modern version of this charades system is what we've already seen for **Chinese characters**. Recall that most characters can be analyzed as containing **two elements**, one of which provides semantic information, while the other provides phonological information, as in the FEMALE and [ma] of "mother." In addition, Chinese uses semantic+semantic composites: the sign for REST consists of PERSON + TREE, the idea being that one can rest under a

tree.

妈

mā "mother"

FEMALE + [ma]

休

xiū "rest"

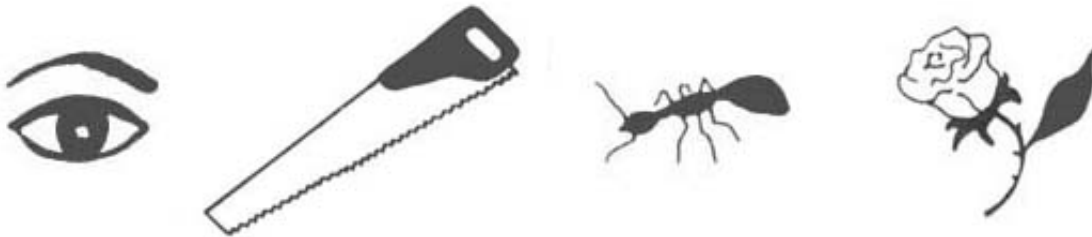
PERSON + TREE

It's important to remember that these "charades" are **completely conventionalized**: an educated reader of Chinese doesn't have to guess every time a familiar character is seen. But when a **rare or unfamiliar** character is encountered, this is precisely the sort of guessing strategy that could reveal the word to which the character refers, without consulting a dictionary.

Rebus



The second Sumerian technique was the **Rebus Principle**: if you can't make a picture of something, use a picture of something with the same (or similar) sound. That is, when a word for some easily drawn object was homophonous with some more abstract word, the originally pictographic sign could be used for the homophonous word as well. A simple English example is "I saw Aunt Rose."



Unlike with Charades, this does not result in a new (composite) sign; rather, it's an **extension** in the use of an existing sign.

Rebus in Sumerian



In Sumerian, a picture of the concrete object *gi* "reed" was used for the abstract verb "reimburse." (Three historical forms of the sign are shown, at 3000, 2400, and 700 BCE; but the rebus extension happened at an early stage and is not related to the change in style over time.)

GI "reed" → GI "reimburse"



The use of this sign for "reimburse" has been called the world's first clear example of rebus, in a tablet from Jemdet Nasr, dated to around 2900 BCE. Rebus extensions solve the problem of writing words for abstract concepts, though they also create a new problem of ambiguity — as in any phonetic writing system that sometimes writes homophonous words identically (cf. English *yard* for a patch of lawn and a measurement).

Similarly, the following sign for a "water" was originally two wavy lines, though in writing with stylus in clay it took on a more abstract appearance. The same sign was also used for the suffix meaning "in" because in Sumerian this suffix had the same pronunciation [a].

A "water" → A "in"

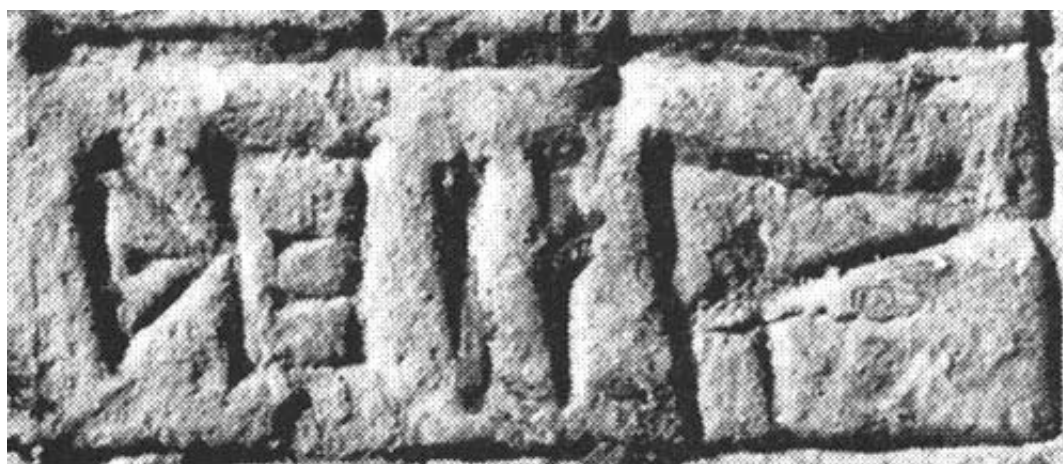


This symbol also soon came to mean the sound [a] in general – regardless of the meaning of the word in which it occurred – thereby acquiring a truly **phonetic** use. For example, it could be used to help write the suffix *ani* meaning "his" or "her." (Chinese uses this technique to write borrowed words: characters are used for their sound, and the meaning is ignored.)

A Sumerian example



The word *nin-ani* "his Lady" was written with three signs: **NIN**, a logogram for "lady"; **A** "water" used phonetically; and **NI** "oil" also used phonetically. Here is that word in a rather archaic style (from around 2100 BCE).



NIN
'lady'

A
('water')

NI
('oil')

Once such phonetic signs become part of the script, it becomes possible to write any word of the language without inventing new logographic signs. The result was a **complete writing system**, in which the Sumerians wrote down not just warehouse records, but poems, diplomatic treaties, letters, contracts and judicial decisions, dictionaries, and epic myths.

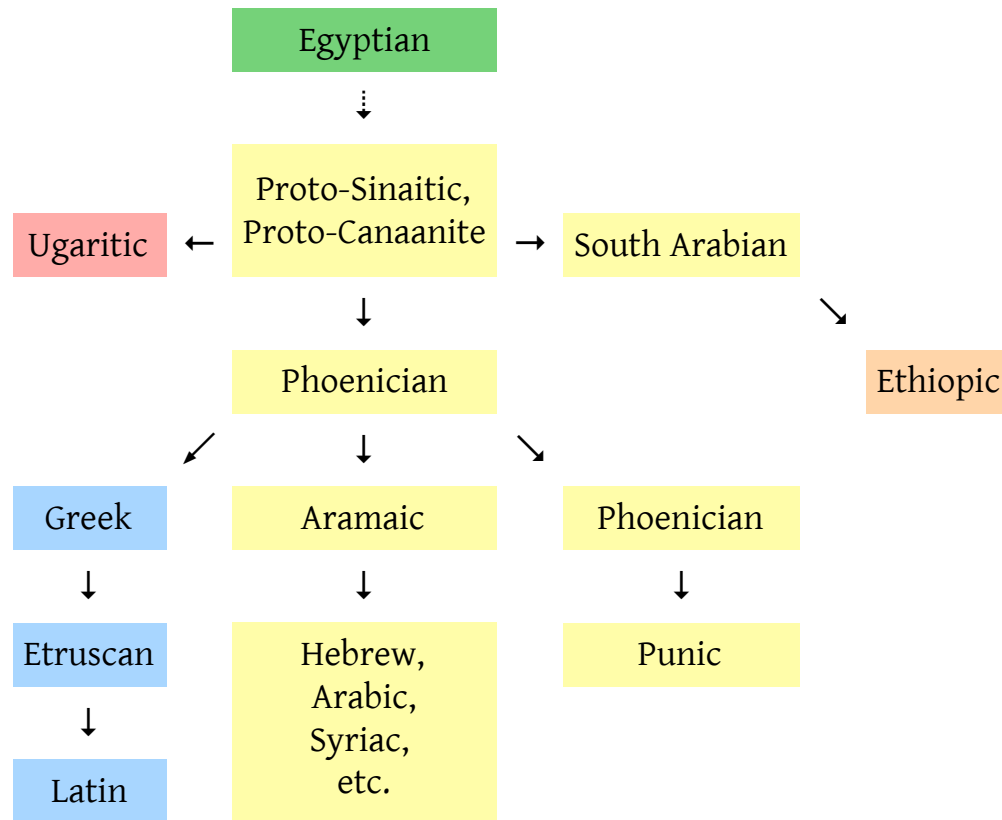
Thus writing seems to have started with pictographs that give only imprecise information about the spoken language equivalents. As the inventory of signs increases, the possibility arises to begin using some of the signs as rebuses or as phonological/semantic combinations. This is much more efficient than trying to design a new symbol for every word or morpheme. Once this meaning-plus-sound process begins, it can develop into a full (if complex and inefficient) writing system, able to encode any utterance in the language.

Developments over time



The **Sumerians** developed a meaning-plus-sound system based on the **syllabic** unit. The Chinese and the Maya did the same. A logical next step is to increase efficiency by doing away with some or all of the meaning-related units, in favor of a consistent syllabary of some sort. Such syllabaries were developed throughout East Asia, but in most cases they did not displace the meaning-plus-sound elements. Instead they supplemented them for certain uses (such as the encoding of grammatical particles in Japanese) or for certain populations (such as women in some places and periods in China).

By contrast, the **Egyptians**, partly due to the structure of their spoken language, developed a meaning-plus-sound system based on **consonants**. This led to purely consonantal writing systems for some of the Semitic languages, notably Phoenician, which led to Hebrew and later Arabic. This chart shows the background of the Latin and related alphabets, which ultimately go back to Egyptian.



It was the Phoenicians who brought the alphabet to the **Greeks**. The innovation of the Greeks was to invent **vowel letters** — but this was essentially by accident, since the Greeks misheard some Phoenician words starting with unfamiliar "guttural" consonants as starting with vowels, and then the first full alphabet, with special symbols for individual vowels and consonants, was born. You can learn a lot more about this in [Ling 115](#), which I teach in the fall.

Same script, different languages



In principle, **any fully linguistic script can be used to write any language**; this is proved by the fact that **foreign names** can be rendered in new scripts. Some scripts are just better adapted to some languages than to others, especially with respect to the individual sounds (in any script) and sound combinations (in a syllabary) that need to be expressed.

The history of writing is filled with examples of scripts adapted to new languages, often unrelated and of very different structures.

DEVELOPED FOR...

USED FOR...

Sumerian

Akkadian, Hittite

Chinese	Japanese, Korean, Vietnamese (formerly)
Arabic	Persian, Urdu, Turkish (formerly)
Hebrew	Yiddish, Ladino
Latin	English, and hundreds more

These examples could easily be multiplied. The adaptations involve changes of varying degrees, such that the result may sometimes be considered a **new script** (such as Greek as adapted to Latin).

Same language, different scripts



Along the same lines, what is essentially the same language can be actively written in different systems. Sometimes this depends on ethnicity or religion, other times on historical shifts.

ONE LANGUAGE...	DIFFERENT SCRIPTS...
Hindi / Urdu	Devanagari / Arabic
Serbian / Croatian	Cyrillic / Latin
Turkish	Arabic / Latin (change in 1928)
Vietnamese	Chinese / Latin (official change in 1910)
South Korean / North Korean	<i>Han'gŭl</i> alphabet with or without Chinese characters
Egyptian / Coptic (Late Egyptian)	Hieroglyphs / Coptic alphabet (from Greek)

Once a writing system is "full", i.e. includes a way to write sounds, it's possible (though not always convenient) to use that system to **write any language**. The point of having a phonetic alphabet such as the IPA is to be able to write down any language at all in the most precise way.

last modified: