Numeration Systems

(March 1, 2001)

Gary Hardegree
Department of Philosophy
University of Massachusetts
Amherst, MA 01003

1. Introduction....1
2. English Number-Words....2
3. Spoken Language *versus* Written Language....3
4. Types of Writing Systems....4
   1. Pictograms....4
   2. Logograms....4
   3. Phonograms....5
   4. Syllabaric Writing Systems....6
   5. Other Syllabaries....8
   6. Phonemic Writing Systems....8
   7. Further Evolution....11
5. Numeral Systems....11
   1. The Egyptian Numeral System....11
   2. Early Greek Numerals....12
   3. Later Greek Numerals....13
   4. Hebrew Numerals....14
   5. Roman Numerals....14
   6. The Babylonian Numeral System....15
   7. The Mayan Numeral System....16
   8. The Hindu-Arabic Numeral System....17
   9. Other Bases....18
6. Back to English Number-Words....19

1. Introduction

As noted in the last chapter "Numbers and Counting", the history of numerical thought seems to proceed as follows. First, we discover numbers, which are discrete quantities. Second, we invent physical tokens (strings, stones, bones, etc.) to represent numbers. Third, we invent words and symbols to represent numbers. This last step presents the *problem of numeration* – how to represent numbers by words and symbols – and a *system of numeration* represents an attempt to solve this problem.

Different cultures have addressed this problem in many different ways. For example, there are quite a few "primitive" languages in which the number-words[1] include only 'one', 'two', and 'many', or even 'one' and 'many'.[2] Most languages, however, have a large variety of number words[3]; for example, English has infinitely-many distinct number-words, as you can readily see by counting and noticing that, no matter how far you count, there will always be at least one more number-word standing at attention

[1] Some authors use the term 'numeral' for any number sign, whether it is a spoken word, a written word, or an ideograph. We reserve the word 'numeral' for the ideographic representation of numbers, and we use the term 'number-word' to refer to spoken words and their phonetic transcriptions, as used to represent numbers.

[2] Check these websites for thousands of number-words from thousands of languages: [http://www.zompist.com/numbers.shtml] [http://euslchan.tripod.com/].

[3] The distinction between "word" and "phrase" is theoretically tricky. For the sake of simplicity, we treat all numerical expressions – including 'one', 'two', 'one hundred', 'five thousand three hundred' – as *words*.

in case you call upon it.[4] We can summarize this amazing fact by saying that the morphology[5] of English generates infinitely-many number-words.[6]

In addition to the ordinary spoken words used to represent numbers (e.g., 'one', 'two' , etc.), many cultures have also invented special graphical symbols to represent numbers, over and above whatever symbols they have otherwise devised to graphically represent spoken words.[7] In some cases, a culture simply takes the initial letter of the spoken word; for example, 'C' is short for 'centum' which in Latin means "hundred"; thus, the Roman numeral for one-hundred. Similarly, 'Δ' (delta) is short for 'deka' which in Greek means "ten"; thus the (early) Greek numeral for ten. In other cases, the numerical symbol is a logogram (ideograph) that represents the number directly; the Hindu-Arabic numerals are examples of this. More about logograms and numerals shortly.

2. English Number-Words

We begin with English number-words, which we learn long before we learn to read or write. The numeration system inherent to English number-words, as well as many other modern languages, is basically a *decimal system*,[8] which means that basic word components arrange into powers of *ten* – 'ten', 'hundred', 'thousand', 'million'. The notable exceptions are the common number-words 'eleven' and 'twelve', which are not decimal.

The English number-word system is a *multiplicative-additive system.* For example, the meaning of the word

two hundred five

is obtained by multiplying 2 times 100, then adding 5 to the result, thus.

two hundred five = 2 times 100, plus 5 = 205

Notice that the order of the word components is critical.

two hundred five ≠ five hundred two
two five hundred = nothing!

[4] In this connection, let us recall John Milton (1608–1674) who said "they also serve who only stand and wait" [from "On his Blindness"; http://www.bartleby.com/101/318.html].
[5] The word 'morphology' is used technically both in Biology and Linguistics; in each case, it pertains to form; in Linguistics, it pertains specifically to the forms and formation of words.
[6] Notice that we don't need fancy words like 'quadrillion' or 'quintillion'. All we need is the principle that 'thousand million' denotes the result of multiplying one-thousand by one-million, that 'million million' denotes the result of multiplying one-million by one-million, etc.
[7] The exception to this are the traditional written languages of China, Korea, and Japan. In Chinese, a written word is an ideograph (ideogram, logograph, logogram), which stands directly for a thing or concept. It bears no resemblance to the spoken word. By contrast, in Western languages, a written word stands for a spoken word, which in turn stands for a thing or concept. There are exceptions – the numerals are ideographs and do not stand for spoken words, but for numbers. This is of course critical to the fact that diverse languages employ the same numerals even when they "pronounce" them quite differently. See Section 4 on writing systems.
[8] From Latin *decima*, which means "tenth".

3. Spoken Language *versus* Written Language

By words, so far I have meant *spoken* words.[9] Spoken language is estimated to be between 50,000 and 100,000 years old, based on paleo-anatomical evidence pertaining to voice boxes of our ancient forebears. However, even before we were articulate *speakers* of language, we probably used a system of *gestural symbols*, and long before that we probably used a variety of *gestural signals*.[10] Eventually, out of these linguistic precursors, spoken language arose, and with it came a momentous change in human culture, which is evidenced in the archaeological records by an explosion of artifacts that are noteworthy both in complexity and variety.[11]

Written language is another matter. It is currently estimated that written language traces back no earlier than the fourth millennium B.C. It is thought to have arisen in Mesopotamia[12] in the Sumerian civilization, who invented a form of writing called 'cuneiform'.[13] Curiously, the language of the Sumerians has no known relatives on this planet! On the other hand, the Sumerians were conquered by a Semitic[14] tribe called the Akkadians; and although the Akkadians did not adopt the Sumerian language, which eventually went extinct, they did adopt much of the Sumerian culture, including *most importantly* their writing system.

Jumping several hundred years ahead, another Semitic tribe – known by the Greeks as the Phoenicians, and known in the Bible as the Canaanites – developed an alphabet, that traces to the Sumerian writing system, and whose linguistic descendants – including most prominently the Greek, Roman, and Cyrillic alphabets – were eventually disseminated world-wide. The Phoenicians were legendary sea-farers and navigators[15]; for example, they settled as far west as Britain, and they circumnavigated Africa.[16] Their largest settlement was Carthage (in present day Tunisia), which is estimated to have had close to a million people at its zenith. But after a number of disastrous wars with Rome,[17] and after the Greeks founded Alexandria, which assumed a central role in Mediterranean sea-trade, the Phoenician civilization declined and then disappeared.

---

[9] In this connection, we note that the fundamental meaning of the word 'language' refers to *spoken* language, and the word itself derives from the Old French word 'langue', which means 'tongue'. The word 'langue' and the word 'linguistics' derives from the Latin 'lingua', which also means 'tongue'. The word 'lingua' appears in the modern English phrase 'lingua franca' which means a *common language*. Originally, 'lingua franca' was an Italian word that referred to a mixed (Creole) language that combined Italian with Provençal, French, Spanish, Arabic, Greek, and Turkish.

[10] In this context, the difference is between a naturally representative gesture (like pointing), which is a *sign* or *signal*, and a purely conventional gesture (like flashing a thumbs-up), which is a *symbol*.

[11] The explanatory idea (theory, if you like) is that a fully-developed spoken language, made possible by a more refined and capable voice box, allowed for vastly more efficient transmission of culture, including its values and technology, from person to person, from village to village, and from generation to generation.

[12] Mesopotamia is a region lying between the Tigris and Euphrates rivers in modern day Iraq. The word 'mesopotamia' is Greek for "between the rivers". For the sake of comparison, consider the word 'hippopotamus' which is Greek for "river horse". This fertile region was home to a number of ancient civilizations – Sumer, Akkad, Assyria, and Babylonia. We will hear from the Babylonians again later.

[13] The word 'cuneiform' is Latin for 'wedge-shaped'.

[14] In Linguistics and Archaeology, the word 'Semitic' refers to a family of languages (and hence tribes) that include Hebrew, Arabic, and Aramaic, to name the most prominent ones. The word 'Semitic' in turn traces to 'Semiticus', which is the Latin name of Shem, who according to the Bible was a son of Noah. On the other hand, the word 'anti-Semitic' refers to a person who discriminates against, or is hostile toward, or is prejudiced against Jews. Ironically, and sadly, the majority of the Semitic tribes of today are anti-Semitic.

[15] This suggests that their astronomical knowledge was very advanced; no satellite-navigation computers in those days!

[16] It has even been suggested that the Phoenicians even made it to the New World, based on certain archaeological findings, but this remains suspect at best. Stay tuned!

[17] These are the Punic Wars, the most legendary of which involved Hannibal and his elephants crossing the Alps. In the recent movie *Gladiator*, the character Maximus is required to "play" a role as one of the "barbarian horde" re-enacting the defeat of Hannibal and the Carthaginians by Scipio Africanus (236?-183? B.C.). But, as the *Gladiator* story goes, in this particular re-enactment, the horde prevails, and Maximus become a hero! Note that the word 'Punic' is simply the Latin

continued…

4. Types of Writing Systems

At this point, it is useful to consider how writing and writing systems are categorized. First, the broad categories of writing are given as follows.

(1) pictograms
(2) logograms
(3) phonograms

1. Pictograms

Pictograms are the earliest form of writing, and employ pictorial characters that *resemble* the objects they represent. The following "dingbats" are examples of pictograms, whose interpretation should be readily apparent.

A pictogram *directly* represents its referent, and is not mediated by any spoken language. Pictograms do not have a pronunciation. Accordingly, pictograms often serve as an "international language"; this is witnessed by the wide-spread use of pictograms known as "icons" in computer software around the world.[18]

2. Logograms

Logograms derive from pictograms, and represent a later stage of writing. Like pictograms, logograms directly represent objects and concepts. Like pictograms, logograms do not have a pronunciation. On the other hand, a logogram does not *pictorially* represent its referent;[19] rather, the connection is *purely conventional*.[20] The following "dingbats" are examples of familiar logograms.

The first three are used on road signs; the latter are used on VCRs, tape decks, and DVD players.

The most extensive system of logograms is the one shared in large part by the written languages of China, Korea, and Japan. In Japan, this system of writing is called ‘kanji’, and the characters are also called ‘kanji’, a word that basically means ‘Chinese characters’. The following are examples of Kanji spatial words.

---

translation for the Greek word ‘Phoenician’, which derives from the Greek word for ‘purple’, which is based on the renowned purple dye invented by the Phoenicians (obtained from a secretion of a sea snail *Murex brandaris*). The modern name of this dye is ‘tyrian purple’ named after Tyre, the capital of Phoenicia (in present day Lebanon). Whereas Tyre was famous for its purple dye, another Phoenician city, Byblos, was famous for its very fine papyrus, which was used for writing (originally by the Egyptians). The word ‘paper’ traces to ‘papyrus’, although paper itself traces to China. The Greeks used the word ‘byblos’ to refer to papyrus, much as we use the word ‘china’ to refer to porcelain originally imported from China. Given the use of papyrus for writing, the word ‘byblos’ later came to mean ‘book’, and eventually this word evolved into the word ‘bible’, which basically means ‘book’.

[18] This is constrained within certain obvious cultural parameters. For example, these pictures probably don’t mean anything to stone age people!

[19] In philosophy and linguistics, the *referent* of a word or phrase is what that word or phrase *refers* to.

[20] Bear in mind that, as with any evolutionary process, there are many "shades of gray" between writing pictograms and writing logograms; some logograms are more pictorially suggestive than others. Also, some logograms are also *phonetically* suggestive, as in Ⓟ. This connects them to the currently predominant form of writing – phonogramic writing.

上 up	下 down	左 left	右 right	前 front	後 back	中 middle

Probably the most widespread logograms are the Hindu-Arabic numerals, which are employed all over the world. These are conventional symbols for the first ten numbers, although the first three are clearly derived from pictograms, indeed the very same three pictograms on which the first three kanji numerals are based.

One	1	一
Two	2	二
Three	3	三
Four	4	四
Five	5	五
Six	6	六
Seven	7	七
Eight	8	八
Nine	9	九

Notice that, whereas the first three numerals are pictorially suggestive of what they mean, the remaining six are purely conventional.

3. Phonograms

The advantage of logograms is that they are efficient idea transmitters; for this reason, they are extensively used for specialized vocabulary – road signs, computer icons, mathematical concepts. The disadvantage of logogramic writing, at least as a *general* method of writing, is that it requires every literate person to learn two different vocabularies, and correspondingly it requires the language community to invent two words for every new object or concept – a written symbol *and* a spoken sound. This uses up a lot of brainpower that *might* be better spent on other matters.

For this reason, in many parts of the world, logogramic writing gradually gave way to phonogramic writing. The basic idea is that pictograms evolved into logograms, which evolved into phonograms.[21] The latter occurred by what is called the "Rebus Principle". A rebus is a picture-word puzzle, such as the following.[22]

[21] This is also occurring nowadays in China, where many logograms perform alternative duty as phonograms, just as happened five thousand years ago in the Fertile Crescent (the crescent-shaped region of alluvial valleys stretching from the Egypt to Mesopotamia).

[22] Hint: they all depict names of TV shows; two of them are current; the remaining one is a classic.

These sorts of puzzles regularly occur in the entertainment section of the newspaper, usually with the comics. Nowadays, most rebus puzzles include some conventional phonograms already built in, as in the first two puzzles. A *pure* rebus puzzle employs only pictorial elements, as in the third example. The Rebus Principle alludes to pure rebus puzzles.

Now, the transformation of a pictogram or logogram into a phonogram basically proceeds as follows.

(1) object *o* is represented in the written language by pictogram/logogram *p*;

(2) object *o* is represented in the spoken language by sound *s*;

(3) the symbol *p* is given an alternative use – to represent the sound *s*.

A very simple example goes as follows. Suppose that our symbol for eye is the following pictogram.

Further suppose that the spoken word for eye is pronounced "º", as in English. By applying the Rebus Principle, we can convert the pictogram 👁, which stands for eye, into a phonogram which stands, not for an object, but for a sound – specifically the sound "º". We can then combine this with other phonograms to represent sound combinations that constitute a variety of words.

Notice that there are a number of variables in the above procedure. For example, suppose we speak Dutch, rather than English; then the spoken word for eye is the sound "½g", and accordingly the pictogram 👁 would become a phonogram for the sound "½g". Along a different tack, suppose that the pictogram 👁 represents, not an eye, but a pea (in its pod). In that case, the corresponding word is the sound "p¶", in English at least, in which case the associated logogram 👁 would stand for the sound "p¶".

4. Syllabaric Writing Systems

There is an important difference between letting 👁 stand for the sound "º", and letting it stand for the sounds "½g" or "p¶", which is evident in the phonetic transcriptions. This corresponds to the difference between *syllabaric*[23] and *phonemic* writing systems.

Most people know what a syl-la-ble is, although it is not easy to give a simple definition. The basic idea is that we can divide a spoken word into phonetic parts, each of which constitutes a simple sound *capable* of being spoken in isolation as an autonomous word. Indeed, many words have syllabic[24] parts that are themselves words; for example, the word ‘themselves’ breaks into ‘them’ and ‘selves’.[25]

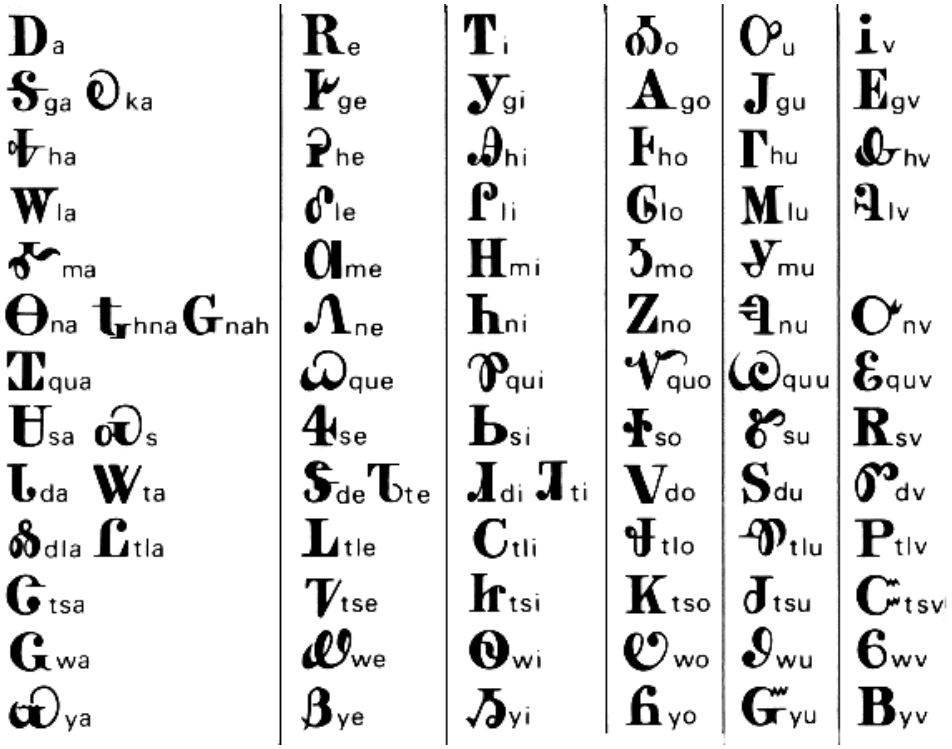
[23] The word ‘syllabaric’ is my own invention, which I propose in place of ‘syllabic’ which is not euphonious.
[24] (s¹-l²b“¹k) – note accent on the middle syllable.
[25] Sometimes the spelling does not cooperate with us – as in the word ‘Northampton’. Notice that the single letter ‘h’ is located in two different syllables, making it impossible to hyphenate this word! One might wonder what happened to the

continued…

In a syllabaric system of writing, each phonogram represents a syllable. In order to construct a syllabaric writing system for a given language, we must compile all the syllables used in that language, and we must assign a unique symbol to each one. The resulting assignment of symbols to syllables is called a 'syllabary',[26] although it is also sometimes called an 'alphabet'.[27]

Probably the most famous syllabary in the U.S. is the one constructed by Sequoyah[28] (1770-1843) for his native language Cherokee. This is probably a unique event in history – a case in which a complete writing system is invented *almost* out of thin air. I say 'almost' because Sequoyah was exposed to the *concept* of phonogramic writing when he enlisted with Andrew Jackson in the War of 1812. He could neither read nor speak English, but he learned that English has a special writing system associated with it, and he learned what the special symbols are. What he did not learn was *any* of the phonetic rules of this writing system. Nevertheless, after several years of work, he managed to phonetically reduce the Cherokee language to 85 syllables, for which he produced his famous syllabary. It is said that, after the introduction of his writing system, the Cherokee nation became literate overnight.[29]

Sequoyah's syllabary for Cherokee is presented as follows. Note carefully that these are lithographic transcriptions of his handwriting (on bark!) to make early typesetting more manageable. Each character is presented next to its Romanic counterpart. Note, in particular, how astonishingly little it shares with English phonetics.



missing 'h' in 'Northampton'. I hypothesize that it was stolen by 'Amherst' [note to non-residents: the 'h' in 'Amherst' is silent, and the accent is on the first syllable].

[26] From which I derive the word 'syllabaric'. See note 23.

[27] The concept is what is important; if we wish, we can distinguish between syllabaric alphabets and phonemic alphabets.

[28] Sequoyah was also known as George Gist, and also known as George Guess. He has many monuments, from Tennessee to Texas, but the biggest and oldest are in California, being the giant redwood trees – sequoia – which are named after him.

[29] Jared Diamond, *Guns, Germs, and Steel: The Fates of Human Societies* (1997; Pulitzer Prize, 1998). From the point of view of cultural history, the transformation in the Cherokee nation did happen "overnight". From another source, one learns that by 1825 much of the Bible and numerous hymns had been translated into Cherokee, and by 1828 Sequoyah's people were publishing the "Cherokee Phoenix", the first national bi-lingual newspaper.

5. Other Syllabaries

There are other syllabaric writing systems, both historical and current. For example, it is widely believed that the earliest Greek writing system was Linear-B, which was used by the Mycenaean Greeks on the island of Crete, and which was not deciphered until 1952.[30] At the other end of history, in the 19th Century in America, a syllabary was invented for the languages of the Cree people (mostly in Canada). This was later adapted for Inuktitut, the language of the Inuit people (Eskimo).

Now, the Mycenaean, Cherokee, Cree, and Inuit nations are not exactly big players on the current world stage. But Japan is! It is the second largest economy in the world today. And they too have developed a syllabaric writing system. Japan currently uses a four-fold writing system, which combines (1) Kanji, (2) Hiragana, (3) Katakana, and (4) Romaji. Whereas Kanji consists of the traditional Chinese logograms, Hiragana and Katakana are syllabaric writing systems, and Romaji is the Romanic writing system. The following is the standard chart.

Table 19-2. Hiragana and Katakana (Japanese Syllabary) in Gojūonzu (50-Sound Chart)

Hiragana	あ	か	さ	た	な	は	ま	や	ら	わ	ん
Katakana	ア	カ	サ	タ	ナ	ハ	マ	ヤ	ラ	ワ	ン
Romaji	a	ka	sa	ta	na	ha	ma	ya	ra	wa	n
	い	き	し	ち	に	ひ	み		り		
	イ	キ	シ	チ	ニ	ヒ	ミ		リ		
	i	ki	shi	chi	ni	hi	mi		ri		
	う	く	す	つ	ぬ	ふ	む	ゆ	る		
	ウ	ク	ス	ツ	ヌ	フ	ム	ユ	ル		
	u	ku	su	tsu	nu	fu	mu	yu	ru		
	え	け	せ	て	ね	へ	め		れ		
	エ	ケ	セ	テ	ネ	ヘ	メ		レ		
	e	ke	se	te	ne	he	me		re		
	お	こ	そ	と	の	ほ	も	よ	ろ	を	
	オ	コ	ソ	ト	ノ	ホ	モ	ヨ	ロ	ヲ	
	o	ko	so	to	no	ho	mo	yo	ro	wo	

6. Phonemic Writing Systems

Syllabaric writing systems are very effective for languages that have a fairly limited number syllables, such as Cherokee and Japanese. There are many languages, however, that have thousands of syllables, which makes syllabaric transcription less than ideal. For example, although Linear-A may have been perfect for transcribing the native Minoan language, its derivative Linear-B proved to be somewhat less than adequate for transcribing Greek. For this reason, the Greeks, and other civilizations, developed what is today the most widespread writing system on our planet – phonemic writing.

Consider the following, understood as spoken words or sounds.

see me ski tea key

[30] That this writing system was a not deciphered for so long explains its mathematical-sounding name. It is an adaptation by the Greek settlers of Linear-A, which was the writing system used by the native Minoans of Crete.

In a syllabaric writing system, each of these sounds (words) would be transcribed by its own symbol. By contrast, in a phonemic writing system, they are transcribed by a combination of symbols based on their phonetic components. For example, although they are quite distinct sounds, all these words share a common phonetic component – they rhyme! Rhyming is not the only lyrical device in speech; there is also *alliteration*, which is illustrated by the following phrases.

big bad bear
hard hearted Hannah
world wide web
million man march

Once again, in each alliterative phrase, although there are three distinct words, they all share a common phonetic component.

Now, the smallest phonetic units (the phonetic "atoms") are called 'phonemes'. For example, each of the above rhyming words can be phonetically decomposed into two phonetic units – a vowel sound (¶), which they all share, and in virtue of which they rhyme – and a unique consonant sound, in virtue of which they are distinct from each other.[31]

Ideally, every phonetic unit is given its own unique symbol. In practice, however, a phonetic system is riddled with compromises, some more objectionable than others. For example, the English phonetic system employs numerous digraphs[32] including the consonant digraphs 'ch', 'th', and 'ph'[33] and the vowel digraphs 'ee', 'oo', 'au', 'ou'.[34] English also employs diphthongs[35] which are sounds that basically combine two smaller sounds – but in novel, although usually predictable, ways.[36] Probably the best example is 'oy', as in 'toy' and 'boy'.[37]

There are no naturally occurring perfect phonetic alphabets. The compromises necessary to use and adapt a writing system "on the run" have left us with numerous imperfect and conflicting phonetic schemes. For this reason, linguists have developed a variety of artificial/planned phonetic writing systems. For example, every dictionary employs a phonetic alphabet. For instance, in the *American Heritage Dictionary*, the word 'degree' is phonetically transcribed as " d¹-gr¶“ ".[38] Other dictionaries for

[31] The word 'ski' actually consists of a vowel phoneme ¶ and a consonant compound, made of 's' and 'k'.
[32] The word 'digraph' comes from Greek and basically means 'two symbols'.
[33] Even this is not entirely satisfactory; there is an obvious difference between 'thatch' and 'that'. Note also that digraphic consonants should not be confused with consonant combinations, such as 'pl' and 'tr', which sub-divide into smaller phonemes (puh+luh, and tuh+ruh). Sometimes we exaggerate the phonemic composition as in "puh-leeeez".
[34] Unfortunately, English writing is a complete disaster from the phonetic viewpoint. For example, the 'oo' is pronounced in at least four different ways – 'door', 'foot', 'poor', 'moot'. This is due to the inter-mingling of conflicting phonetic systems on the British Isles, as many cultures took control of various parts – Celts, Angles, Saxons, Vikings, Normans. For example, whereas the pronunciation of 'door' is exactly like Dutch; the pronunciation of 'moot' is exactly like French (although it occasionally seems that no one outside France can actually pronounce French correctly!) Oftentimes, a vowel discrepancy reflects a difference between an Anglo-Saxon pronunciation, as in 'mouth' (a common word, spoken by commoners) and a Norman pronunciation, as in 'uncouth' (an uncommon word used to describe commoners).
[35] 'diphthong' (d¹f“thông", -th¼ng", d¹p“-), literally means 'two tongues' [Middle English *diptonge*, from Old French *diptongue*, from Late Latin *dipthongus*, from Greek *diphthongos* : *di*, two + *phthongos*, sound.].
[36] Not all diphthongs are represented by vowel pairs in English; the 'i' in 'fine', and the 'a' in 'face' are diphthongs. Note, however, that many upper-class Brits, and some of their linguistic-relatives in the Southern U.S., do not pronounce these as diphthongs.
[37] As with most children, my daughter had a number of charming words when she was young. For example, she pronounced 'toys' like 'toe ease' and 'boys' like 'bow ease'. Eventually, however, this diphthong was mastered.
[38] I completely disagree with this assessment, but I am no authority! I believe it is pronounced " d…-gr¶“ " as in 'Hardegree'!

other languages employ other phonetic schemes inasmuch as each language has its own phonetic idiosyncrasies.[39]

For this reason, various more ecumenical phonetic alphabets have been devised, including the International Phonetic Alphabet.[40] The chief goal of this endeavor is to transcribe the huge variety of languages on our planet, of which there are currently six-thousand or so. This task is vitally important, not only to theoretical linguistics (phonetics and phonology), but also to anthropology. This is because hundreds of languages on our planet are endangered, and many go extinct every year. Since these cultures have no independent written record of their native language, the only way to preserve their languages for subsequent study is to transcribe them before the native speakers die off.

The International Phonetic Alphabet has a number modules, probably the simplest of which pertains to pulmonic consonants (the ones involving our pulmonary system – i.e., lungs).

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Retroflex
Plosive	p b			t d		ʈ ɖ
Nasal	m	ɱ		n		ɳ
Trill	ʙ			r		
Tap or Flap				ɾ		ɽ
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ
Lateral fricative				ɬ ɮ		
Approximant		ʋ		ɹ		ɻ
Lateral approximant				l		ɭ

If nothing else, the *vocabulary* of phonology and phonetics can be a bit daunting. Nevertheless, many of the terms make sense once you realize what parts of your body are used to make these sounds. For example, the "dental" consonants are made using the teeth, and include the two ‘th’ sounds, as in ‘thatch’ and ‘that’.

[39] For example, the Cyrillic alphabet is an adaptation of the Greek alphabet for the purpose to transcribing Slavic languages. Notice that, whereas Greek as 24 letters, Cyrillic has 35 letters. It is named after Saint Cyril (827-869), who was an Eastern Orthodox missionary to the Slavic peoples of Moravia and Russia. Saint Cyril’s linguistic task was to compile the phonemes of the native languages and figure out which ones could be transcribed into Greek, and to propose new symbols for the phonemes that could not be so transcribed. Not all Slavic speaking peoples employ the Cyrillic alphabet, however. By and large, whereas Eastern Orthodox nations (e.g., Russia) use the Cyrillic alphabet, Roman Catholic nations (e.g., Poland) use the Roman alphabet. This is not a coincidence!

[40] Consult their website: [http://www2.arts.gla.ac.uk/IPA/ipa.html].

7. Further Evolution

As we have described it so far, in the ordinary course of events, pictograms evolve into logograms, and logograms evolve into phonograms. However, it is also possible for phonograms to evolve into logograms. Probably the most colorful example is the ampersand symbol '&' which has mostly shed its phonogramic ancestry,[41] and now simply means 'and'. But ampersand evolved from the Latin word 'et', which is a phonogram. The symbol was evidently invented in 63 BC by the Roman Tiro. In the 19th Century, English school children were taught an alphabet with 27 characters, including the symbol '&'. Three of these letters are also used as words – 'a', 'i', and '&' . To distinguish the words from the letters, the words were called 'a *per se* a', 'i *per se* i', and 'and *per se* and'. The Latin '*per se*' means 'in, or by, itself'. Evidently, English school children were unable to master the pronunciation exactly, and we inherited the word 'ampersand', which is the name of the symbol '&'.

Some fonts render ampersand in a manner that is highly suggestive of its Latin origins, but others render it more abstractly. The following is a sample of true-type font renderings of ampersand.

& & & & &

There are also words that are so phonetically isolated that they might as well be logograms – for example:

one two

In the word 'one', the 'on' is phonetically recognizable; it is reminiscent of 'son', 'money' and 'London'. The 'e' does *no* work, and the 'w' sound appears *ex-nihilo*. In the word 'two' the 't' is recognizable, but the 'w' – which was obviously stolen from 'one' – plays no phonetic role. The 'o' is recognizable, being reminiscent of 'do' and 'to'.

Finally, we note that it is entirely possible for a phonogram to evolve into a logogram, which in turn evolves into a completely different phonogram. See if you can read the following syllabaric rebus puzzle.

m&8

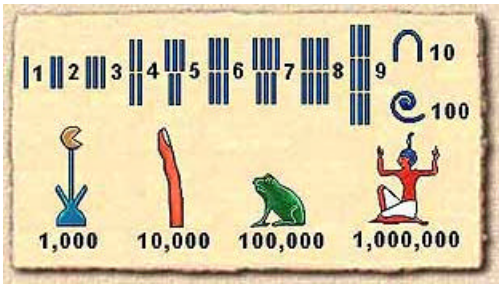
5. Numeral Systems

As mentioned earlier, various cultures have invented special symbols for numbers, over and above the symbols provided by spoken language. It is customary to call these special symbols 'numerals'. In what follows, we will examine a few of these numeral systems.

1. The Egyptian Numeral System

One of the earliest examples of a numeral system is the Egyptian numeral system, based on the following hieroglyphs[42].

[41] Note, however, that '&c' is still read 'et cetera'.
[42] The word 'hieroglyph' combines 'hiero' [holy] and 'glyph' [carving].



Notice that the first nine numerals are pictographic in character, but the remaining ones are logographic in character.[43] Notice also that this is a decimal system. What the above illustration doesn't tell us, however, is how the Egyptians wrote compound numerals. As it turns out, the Egyptians used a simple *additive system*, as illustrated in the following diagrams.[44]

Note carefully that although the Egyptian numeral system does not especially *require* a symbol for zero, the Egyptians nevertheless *had* a symbol for zero

which they used for a variety of engineering and accounting purposes, including some rather astonishing projects, such as the Pyramids (c. 2550 BC).

2. Early Greek Numerals

There are two sets of ancient Greek numerals. The *early* Greek numeral system (Attica, c. 1000 BC)[45], employed the following atomic numerals.[46]

I	Γ	Δ	Γ′	H	Γ″	X	Γ‴	M
1	5	10	50	100	500	1000	5000	10000

Whereas the character 'I' (iota) is a basically a pictogram for the number one, common to many cultures, the remaining characters are phonogram-derived. For example, the symbol 'Γ' is an early version of 'Π' (pi), which is the first letter of 'penta', which means 'five'. Similarly, the symbol 'Δ' is delta, which is the first letter of 'deka' which means 'ten', and the symbol 'H' is eta, which is the first letter of 'hekaton', which means 'hundred'.

---

[43] The symbol for 1000 is a picture of a papyrus plant, which is very important to Egyptian civilization and to civilization in general. The word 'paper' derives from the word 'papyrus' which refers to the dominant writing medium for many years in the Mediterranean civilizations. It was eventually superseded by paper, which is made from wood pulp, by a process invented in China and brought back to the West by the renowned Italian explorer Marco Polo.

[44] Actually, the Egyptians wrote from right to left, as in Arabic and Hebrew. But, since it is a simple additive system, the order doesn't really matter!

[45] Attica is the region surrounding Athens.

[46] The symbol 'Γ' is not a gamma but only our rendition of an early version of the Greek letter Pi, for 'penta'. Similarly, the various primed "gammas" are rather poor renditions of the Greek decorations to convey these numbers.

Like the Egyptian system, the early Greek system is a simple *additive system.* For example, to read the numeral

ΔΔΓIII

one simply adds the values of the separate numerals. Thus:

ΔΔΓIII = 10 + 10 + 5 + 1 + 1 + 1 = 28

Notice in particular that the order of the terms is irrelevant to the eventual calculation of the number represented. Nevertheless, the conventional ordering is in terms of descending value left to right.

3. Later Greek Numerals

The numeral system of Attica was eventually superseded by the numeral system of Ionia, an early Greek settlement in western Asia Minor (modern day Turkey). The Ionian numeral system employs the standard classical Greek alphabet, along with three archaic letters. The code is fairly straightforward.[47]

1	α	10	ι	100	ρ
2	β	20	κ	200	σ
3	γ	30	λ	300	τ
4	δ	40	μ	400	υ
5	ε	50	ν	500	φ
6		60	ξ	600	χ
7	ζ	70	ο	700	ψ
8	η	80	π	800	ω
9	θ	90		900	

The characters for 6, 90, and 900 are respectively digamma, koppa, and sampi, which are early Greek letters that fell into disuse as phonograms, but continued to be used as numerical logograms.

The system is evidently a decimal system. Although it isn't evident from the chart, the Ionian numeral system is also a simple additive system, like the Egyptian and Attic systems. For example:

φμβ = 500 + 40 + 2 = 542

[47] This code is the basis of the "Bible Code", which has exercised many people over the years. Note that the earliest New Testament was written in Greek, and indeed the name 'Jesus' ('ιησους') is Greek, probably being a transliteration of 'Joshua'. Similarly, the name 'Christ' ('χριστος') is Greek for 'anointed'. Note in this connection that, prior to the Roman annexation of Palestine by Augustus Caesar, the Holy Land was under Greek rule, under a dynasty tracing to Alexander the Great, who took it from the Persians, who took it from the Babylonians, who took it from … But, back to the code: each letter in 'ιησους' is also a numeral [note: 'ς' is the word-ending variant of 'σ'], and each numeral stands for a number, and when you add up these numbers, you get 888. This is fascinating, to be sure, but its precise religious significance is way beyond my power to fathom.

4. Hebrew Numerals

Alexander the Great conquered Persia, the Middle East, and Egypt during the 4th Century BC. One of the lands he conquered was Judah, the homeland of the Hebrew people. Sometime during the 1st Century BC, the Hebrews transcribed the Ionic numeral system into their alphabet, thus producing the following numerical code.[48]

1	à	10	é	100	÷
2	á	20	ê	200	ø
3	â	30	ì	300	ù
4	ã	40	í	400	ú
5	ä	50	ð	500	–
6	å	60	ñ	600	–
7	æ	70	ò	700	–
8	ç	80	ô	800	–
9	è	90	ö	900	–

Thus, just as with the Ionic alphabet, the Hebrew alphabet provides *both* phonograms for transcribing spoken Hebrew, and logograms for representing numbers.

5. Roman Numerals

Before Rome, the most developed civilization on the Italic Peninsula[49] was the Etruscan civilization, who copied their numerals from the early Greek (Attic) system. These in turn were adopted and adapted by the Romans, who formulated the Roman numeral system, still in wide use today for a variety of purposes. As every grade school child can tell you, the Roman numeral system is based on the following seven atomic numerals.[50]

I	V	X	L	C	D	M
1	5	10	50	100	500	1000

The Roman numeral system is not a simple additive system, but is rather an *additive-subtractive system.* Indeed, the subtractive aspect is frequently a source of consternation when reading large numerals – for example:

MCMXCIX

[48] This too is a source of numerological mysticism in the Jewish faith.

[49] So-called because it is slanted! No, not really. The slanted writing system that we call 'italics' was developed in Italy during the Renaissance, and was accordingly called 'italic'.

[50] There are other Roman numerals that most of us never learn, but can be found in Latin dictionaries – for example:

5000	Iↄↄ
10000	ↀↀIↄↄ
50000	Iↄↄↄ
100000	ↀↀↀIↄↄↄ
500000	Iↄↄↄↄ
1000000	ↀↀↀↀIↄↄↄↄ

etc.

By saying that the Roman system is (partly) subtractive, we mean that some combinations of symbols require us to apply subtraction in order to interpret them. For example,

IV

stands for "one before five", which is four [i.e., 5 *minus* 1]. Similarly, the numeral

XC

stands for "ten before one-hundred", which is ninety [i.e., 100 *minus* 10]. On the other hand, the string

IC

is officially *ill-formed*, although it *could* be understood to mean "one before one-hundred", which would then be ninety-nine.

So how do we interpret a Roman numeral such as 'MCMXCIX'?

M	is not before a larger numeral, so it reads:	+ 1000	1000
C	is before a larger numeral, so it reads:	− 100	
M	is after a negative prefix, so it reads:	+ 1000	900
X	is before a larger numeral, so it reads:	− 10	
C	is after a negative prefix, so it reads:	+ 100	90
I	is before a larger numeral, so it reads:	− 1	
X	is after a negative prefix, so it reads:	+ 10	9

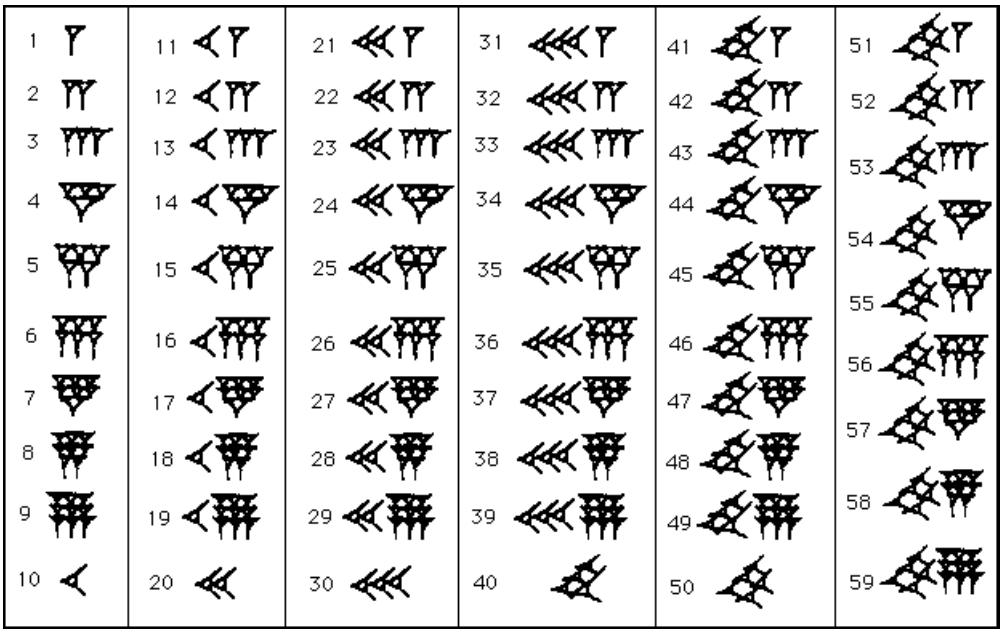
Thus, 'MCMXCIX' represents the number 1999.

6. The Babylonian Numeral System

The systems we have discussed so far are characterized by the fact that every instance of a given atomic numeral or number-word has the same meaning no matter where it occurs. 'I' means "one" wherever it occurs; 'X' means "ten" wherever it occurs; and so forth. The numeral system of the Mesopotamian civilization of Babylonia presents a radical departure from this scheme. Specifically, the Babylonians employed an incredibly sparse numeral system, which employs just two cuneiform symbols (borrowed from the Sumerians).

The first symbol stands for the number 1, and the second one stands for the number 10 (usually! see below). The following are examples of Babylonian numerals.[51]

[51] From [http://www-groups.dcs.st-and.ac.uk/~history/HistTopics/Babylonian_numerals.html].



Note that the Babylonian system is the first example of a *place-value system*, which we will discuss further in Section 8, when we discuss Hindu-Arabic numeration. In particular, in addition to meaning 1, the symbol 𒁹 also means 60, 3600, 21600, …, as well as 1/60, 1/3600, 1/21600, … In other words, the Babylonian system is a base-60 (sexagesimal) system.[52] It is worth noting that, although the Babylonian numeral system is a base-60 system, it only has 59 basic numerals, since it lacks a symbol for zero! However, given how large the base is, the need for a symbol for zero was comparatively rare, so this problem was not very serious.

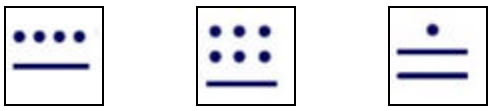
7. The Mayan Numeral System

So far in our discussion of numeral systems, we have only examined cultures that form a direct part the European cultural heritage. In this section, we very briefly digress to consider a completely independent culture – the Mayans. The Mayan civilization was formed as early as 1500 BC in Mesoamerica[53], and reached its zenith between 300 and 900 AD, during which time they built the structures for which they are remembered, including some remarkable temples and pyramids.[54] The numeral system they developed was a lot like the Babylonian system, the difference being that:

(1) the Mayans used three basic symbols:
a dot for 1;
a dash for 5;
a special picture symbol, of a nut or seed, for zero;
(2) the Mayans used a base-twenty (vigesimal) scheme.

The following are a sample of Mayan inscriptions, whose interpretation I leave to the reader.

[52] In this connection, notice that our culture has adopted a sexagesimal system in connection with time-keeping and the measurement of angles. There are sixty minutes in an hour, and sixty seconds in an hour. These very same words are used for angular measurement. A degree of angle or arc is divided into sixty minutes, each of which is divided into sixty seconds.
[53] Mesoamerica (literally "middle America") is regarded to be a region extending south and east from central Mexico to include the Yucatan peninsula, as well as parts of Guatemala, Belize, Honduras, and Nicaragua.
[54] There are hare-brained theories intended to account for the mysterious "coincidence" of pyramids in Mesoamerica and Egypt; they fail to note the 3000 year time difference!



8. The Hindu-Arabic Numeral System

Having made a brief detour, we now continue discussing numeration systems leading up to our own. The next big breakthrough in numerical science and technology occurred in India some time in the 6th Century AD, at a time when Europe was in the midst of the "Dark Ages". This knowledge in turn was passed on to the West in the 8th Century AD, *via* Baghdad, a thriving commercial center (not far from the ruins of Babylon), and finally reached Europe in 1202 AD, when Leonardo of Pisa (a.k.a. Fibonacci[55]) published his *Book of the Abacus.*[56]

The specific breakthrough in India was to combine the place-value scheme of the Babylonians with the decimal scheme of the (later) Greeks. A place-value system is so-called because, in such a system, a given atomic symbol does not have a unique meaning (value) *once and for all*, but has a meaning *dependent upon its location* (place) within the overall symbol.[57]

In the Hindu-Arabic numeration system, with which we are all familiar, we begin with ten primitive (atomic) numerals

0, 1, 2, 3, 4, 5, 6, 7, 8, 9

and construct all the compound numerals by simple concatenation.

There are three key ideas involved in this construction.

(1) **Place Value**: the numeral '1' does not stand *univocally* for the number one, but rather stands *ambiguously* for infinitely-many powers of ten – one, ten, hundred, thousand, …[58] That these numbers are powers of ten is precisely the sense in which this system is a *base-ten* system. How we understand a given atomic numeral, in particular, depends upon its location in the overall molecular numeral. For example, in '125', '1' means one-hundred, but in '17', '1' means ten.

(2) **Additivity**: the numbers represented by the atomic numerals are simply added together to evaluate a molecular numeral. Thus, for example,

'125' stands for one-hundred + twenty + five = one-hundred twenty-five

[55] Fibonacci grew up in Northern Africa, where his father was a diplomat, and where he learned of the new numeration scheme. Check here: [http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Fibonacci.html] [http://www.mcs.surrey.ac.uk/Personal/R.Knott/Fibonacci/fibBio.html].

[56] The novel numeral system did meet some political and legal resistance. For example, in Florence the Arabic numerals were outlawed, because they were much easier to counterfeit and alter than Roman numerals, and accordingly posed a serious threat to commerce and banking. This distrust of Hindu-Arabic numerals is still present today; for example, no check is acceptable unless it has the amount written as an English number-word.

[57] Note also that the context-dependence of a numerical valuation is also present in the Roman numerals, insofar as 'XI' and 'IX' don't mean the same thing: 'I' before 'X' means subtraction; 'I' after 'X' means addition. Similar context-dependence is also present in the English number-words: 'two' before 'hundred' calls for multiplication; 'two' after 'hundred' calls for addition.

[58] Not to mention all the negative powers (fractions) – one/tenth, one/hundredth, etc.

(3) **A Symbol for Zero**: this idea took a little while to develop within the Hindu numeral system; it is the recognition that, in a place-value system, a numeral for zero is required in order to position the digits precisely.[59]

The third feature of Hindu-Arabic numeration is sometimes referred to as the "discovery of zero", but this is absurd for a couple of reasons. First, the Egyptians had a symbol for zero three millennia before this, although it was not necessitated by their numeral system, which was a simple additive system. Second, the *number* zero was probably discovered almost as soon as numbers were discovered. 'Zero' is after all synonymous with 'none'. What took thousands of years was to invent a special symbol for the number zero, and then to incorporate it into a system of numeration.

9. Other Bases

Most elementary school students learn that the base-ten numeral system does not enjoy a monopoly in the world of ideas, but is one among many bases in use. The most important alternative bases are base-two (binary), base-eight (octal), and base-sixteen (hexadecimal), because of their prominent roles in computer science.

Whereas the binary system is vitally important to the design and function of computer hardware and machine-language software, the hexadecimal system is important in high-level programming languages (Basic, Pascal, C++). The reason is that information is often stored in "ASCII" format, which is to say it is stored as a series of bytes. Each byte consists of eight bits, and each bit admits two values (0 and 1). Accordingly, each byte has 256 ($=2^8$) values , each of which can be represented by a two-digit hexadecimal numeral. The hexadecimal numeral system is based on the following atomic numerals.

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
										10	11	12	13	14	15

As with every place-value system, the numeral '1' stands for a different number according to its location within the compound numeral. In the case of the hexadecimal system, the possibilities include all the powers of sixteen:

one
sixteen
two-hundred fifty-six
four-thousand ninety-six
sixty-five thousand five-hundred thirty-six
…

As with all numerals within a place-value system – be it Babylonian, Mayan, Hindu, or "Hexinian" – a hexadecimal numeral is evaluated using powers of its base (radix), which for Hexinian is sixteen. For example,

FF	=	15×16	+	15	=	255
AA	=	10×16	+	10	=	160
11	=	1×16	+	1	=	17

[59] Note that the "problem of zero" is not a problem with simple additive systems.

Here, we have performed the calculations using the familiar base-ten numerals. By contrast, the following are examples of "native" hexinian calculations.

```
  34      77      88
+ 23    + 88    + 99
  57      FF     121
```

Notice that the first one *looks* perfectly ok, but its *meaning* is completely different from its meaning in "decimalese", since in Hexinian, ‘34’ means fifty-two, ‘23’ means thirty-five, and ‘57’ means eighty-seven.

6. Back to English Number-Words

We briefly return to English number-words to compare them with the Hindu-Arabic numeral system. Both are decimal systems, but they are not both base-ten systems. In particular, whereas the Hindu-Arabic numeral system is a place-value system, with radix ten, the English number-word system is not. The former requires a symbol for zero; the latter can live without it. This is emphasized by the fact that when we translate a Hindu-Arabic *numeral* into English, we often elide[60] one or more zeros. For example, we read the following numerals thus.

1<u>0</u>5	one hundred five
1,<u>00</u>5	one thousand five
1,<u>000</u>,500	one million, five hundred

Notice the unpronounced zeros.

Finally, we note an interesting usage fact. In particular, we note the common practice of simply reading the numeral ‘105’ as "one-oh-five". This is not remarkable in itself. But there are interesting usage differences between ‘one-oh-five’ and ‘one hundred (and) five’. The former is perfectly fine for ordinal numbers (e.g., room numbers), but is not entirely acceptable as an adjectival-word. For example, the sentence

> there are one-oh-five members in this club

seems very odd, if not ungrammatical Conversely, it sounds equally odd (even ungrammatical) to give one’s address as

> one hundred and five Main Street,

or to advise someone in trouble to dial

> nine hundred and eleven.

[60] *elide* – To omit or slur over (a syllable, for example) in pronunciation. This word also refers to the syntactic process, called ‘ellipsis’ (≠ ‘ellipse’!), which consists of omitting phrases as in “I don’t have a pen; do you …?” Here, we have elided ‘have a pen’. Incidentally, the “dot, dot, dot” notation is called ‘an ellipsis’.