The digital origin of human language—a synthesis

Hans Noll

Summary

The fact that all languages known are digital poses the question of their origin. The answer developed here treats language as the interface of information theory and molecular development by showing previously unrecognized isomorphisms between the analog and digital features of language and life at the molecular level. Human language is a special case of signal transduction and hence is subject to the coding aspects of Shannon's theorems and the analog aspects of pattern recognition, each represented by genotype and phenotype. Digital language acquisition is late in evolution and postnatal development and requires a neural reorganization by a mechanism of somatic network programming in response to the environment. Such a mechanism would solve the Chomsky conundrum of how children can learn any language without knowing rules of grammar too numerous to be encoded genotypically. *BioEssays* 25:489–500, 2003. © 2003 Wiley Periodicals, Inc.

> "To create is to recombine"
>
> François Jacob

Introduction

The two components of information: digital encoding and analog interaction of patterns

The fact that all human languages make use of a digital code has not received general attention despite its crucial importance for an understanding of the origin of human language and consciousness. Here I propose that the digital nature of human language is a late development in the evolution of *Homo sapiens* and the key event that separates humans from the other hominids. Let me begin with an explanation of the terms digital and analog for readers not familiar with information theory. Digital refers to a method of encoding information by a linear sequence of elements or signs chosen from a limited set of numbers, letters or sounds. The combinatorial principle makes it possible to produce an infinite number of different sequences, regardless of whether we use a set of two elements as in the binary code of the digital computer and the Morse code, or a set of four as in DNA or a set of between 20 and some 50 elements as in the amino acid sequences of proteins and the alphabets of most written languages. Most of these codes are logically equivalent and are interconvertible by a simple rule or algorithm. The basic principle is that in digital communication the code is independent of meaning, and hence the same code can be used to express a variety of different languages. By contrast, Chinese script is not representing a phonetic, and thus digital, alphabet, but is based on pattern recognition in which each sign is associated with a specific meaning and hence independent of the spoken language. This aspect is illustrated by the fact that people speaking the multitude of different dialects in China can communicate by means of the universal analog vehicle of Chinese writing if they have been taught to recognize the pattern and meaning of at least 3,000 different signs required for the vocabulary of every-day language.

While the structure of all known languages is clearly digital by its subdivision into syllables, words and sentences and their combinatorial arrangement into linear sequences, important aspects of the transmission, i.e., the acoustic generation and perception of language as well as its meaning are analog events and thus based on pattern recognition. This distinction is important because most people indiscriminately associate information with meaning and digital encoding. As Claude Shannon, the father of information theory, pointed out, a rigorous, mathematical definition of information is only possible, if it is purely a matter of coding and the physics of message transmission and thus totally dissociated from what we understand as meaning or content.(1) For this reason, Gell-Mann(2) and others associated with the Santa Fe Institute have proposed the term complexity for information in the every-day sense of meaning. So far all attempts to describe complexity mathematically have failed. While complexity is the controlling feature of all evolving systems, I prefer the functional definition of "pattern recognition"(3) or "productive interaction of patterns", or in short, "analog", to describe that aspect of information commonly referred to as meaning or content.[1]

American Cancer Society Professor of Genetics and Molecular Biology, 6716 36th Avenue NW, Seattle, WA 98117.
E-mail: hansnoll1@attbi.com

DOI 10.1002/bies.10281

---

[1] My definition of analog includes all aspects of information that are not digital in the quantifiable Shannon sense. That comprises all sensory inputs as they are perceived as visual or auditory signals in communication. Analog in this sense is a convenient substitute for "pattern recognition", a term first proposed by Bresch(3) to define the non-Shannon aspects of language, or, more precisely, "productive interaction of patterns," which conveys the notion that patterns must interact with other patterns in communication or signal transduction. John von Neumann was probably the first to use it in this extended sense in his discussion of brain function.(44)

Shannon and the transmission of information

If Planck and Einstein revolutionized our view of the physical universe in the first half of the 20th century, it was Shannon who initiated the information age that came to dominate the second half. His quantitative mathematical definition of information as a statistical problem of encoding devoid of meaning is both bold and deep and transformed mankind more than any discovery since the advent of writing. The basis of a trillion dollar industry is Shannon's mathematical determination of the channel capacity, the rate in bytes per second at which information can be transmitted. He further showed mathematically that the distorting influence of noise encountered during all message transmissions may be minimized by addition of redundancy. His quantitative definition of informtion turns out to be formally analogous to that of entropy in themodynamics. It is important, however, to point out that Shannon entropy, as it is often called, applies to a different phenomenon than thermodynamic entropy, as reflected in the different units of measurement. As Planck(4) wrote: "Since the entropy S is an additive magnitude but the probability W is a multiplicative one, I simply postulated that $S = k \log W$, where k is a constant," and W is a measure of molecular disorder. Shannon argued similarly that it was convenient to be able to add linearly the probability units of information and defined as unit the probability of choice between 0 and 1 in the binary system, or one "bit". N bits of information would thus correspond to $\log_2 2^N = N$. Applied to a language encoded by 32 letters, one letter would correspond to 5 bits ($\log_2 32 = 5$). Since the number of signs in word processing of western languages is about 100, 7 bits (= 128 alternatives) would be required per letter. However, most modern computers encode 8 bits = 1 Byte. This ensures sufficient redundancy for protection against noise in transmission. This last point is an important one: redundancy of information helps promote accuracy of signal reception.

When Shannon(5) analyzed the structure of the English language statistically, determining the relative frequency of all letters, letter combinations and word sequences, he found that spoken language was about 50% redundant, a fact that greatly facilitates understanding. Here, Shannon's theory clearly explains and is confirmed by evolution, a fact encountered again in the discussion of the genetic code. Shannon, it should be recalled, did his Ph.D. work at the Cold Spring Harbor Laboratory on population genetics.(6) An excellent illustration of Shannon's redundancy prediction is the popular television game "wheel of fortune". The best chances of winning have those participants who select the letters according to the Shannon frequencies. Experience shows that the message is usually read correctly when 30% to 50% of the letters are still missing.

The distinction between Shannon-information in the form of digital encoding on the one hand and information as meaning on the other, is best illustrated by the problems facing cryptology. Here a message containing a specific meaning is to be digitally scrambled, by a computer program or encoding machine, to make it unintelligible to any recipient without a key to restore the original order. The message is only safe as long as the code is not broken or the coding machine or procedure not captured by the enemy, as occurred repeatedly during WWII. Another method is to translate the message into a foreign language resistant to digital decoding methods. When the US used for this purpose the Navajo language, which was unknown to the Japanese, this code was never broken because the key is not a digital procedure but the pattern recognition function of a brain knowing both languages.

The chasm between digital encoding in spoken and printed communication on one hand, and the analog meaning of the content on the other, often gives rise to conflict because comprehension must occur at two levels: it is not enough that the recipient speaks the same digital language, he must also share the speaker's pattern recognition as illustrated by Thurber's fable of the man who tells his wife that he saw a unicorn in the garden eating roses.(7) After telling him that the unicorn is a mythical beast, she calls the police and a psychiatrist with a straitjacket. Listening to the wife, the police and the psychiatrist seize her and put her into the straitjacket. Asked by the police upon his return from the garden, whether he told his wife that he saw a unicorn, the husband replied: "Of course not, the unicorn is a mythical beast." Satisfied, the psychiatrist told the husband that his wife was "as crazy as a jay bird" and ordered the police to take her away to an institution. Another illustration is what occurred when, on a visit to friends, our 4-year old son, Lucas, was playing outside with Sally, also 4, of our guests. At one point, Sally came storming into the house crying: "Lucas pinched me! " Her mother: "where? " Sally: "on the porch." Less amusing are the conflicts resulting from the actions of politicians with incompatible perceptions of the same reality.

Importance of brain size

Communication by sound offers many advantages and hence its selective value in evolution is obvious, especially for social animals. Animals are masters in recognizing visual as well as acoustic patterns, an essential condition for survival. Birds are probably the most sophisticated acoustic communicators and some species surpass even man in the production and recognition of sound patterns. Mina birds, for example, are capable of reproducing not only any melody precisely but also the characteristics and timbre of individual human speech or laughter. Recent experiments imply that gray parrots even have a limited understanding of human language. This raises the question of why such birds have not developed a more extensive analog or semantic vocabulary, since neither the production and recognition of sound patterns, nor memory appear to be limiting. The limiting factors must be brain size and the social conditions of prolonged proximity required for learning an extended vocabulary, a slow process requiring two to three years in children with a much larger brain than birds.

In the evolution of hominids, we would expect brain size, tool making and phonetic analog communication to have evolved together and in a mutually dependent fashion. Yet, contrary to this expectation, during the time of the greatest brain growth beginning about 2 million years ago, there is no corresponding progress in tool making.(8) Nor is there any other evidence for the existence of a modern-type language even though the anatomy for producing modern speech also evolved during this period and must have been perfected long before modern *H. sapiens* appeared on the scene during the middle and upper paleolithic some 100,000 to 30,000 years ago. The force driving the rapid growth of brain size is, according to Dunbar, the selective advantage of living in larger groups which requires more sophisticated communicative skills than hunting and tool making.(9) According to this view, the information exchanged was not the factual type but rather the emotional sort used for bonding, similar to what is known today as cooing and grooming, e.g., in baby talk, the interminable telephone chatter of teenagers or the way people talk to their pets.

The sudden and simultaneous appearance of more sophisticated tools, symbolic engravings, and cave paintings about 30,000 years ago is a dramatic break with the past and coincides with the arrival of *H. sapiens sapiens* known as Cro-Magnon in Western Europe. The explosive nature of this revolution is consistent with the perfection of a digital language, a conclusion also corroborated by the observation that in children the faculty of drawing pictures from nature with a fair degree of verisimilitude is a late development arriving not before age 8 or 9, several years after reading and writing.(10,11) Inexorably, the digital principle, once discovered and integrated with the analog power of pattern recognition, opened up vast new possibilities in the realm of thinking, cognition and consciousness, and, because of its inherent ease of transmission, would spread rapidly. The advantages gained were so enormous that the people who acquired this superior means of communication, quickly dominated those lacking it or unable to adapt. The disappearance of the Neanderthals, after some 100,000 years of coexistence with *H. sapiens*, has been attributed to this new development, a conclusion subject to doubt.(12)

The rapid spread, universal adoption and exclusive survival of a digital phonetic language has its parallel in the emergence of an universal genetic code since the inception of life more than three billion years ago.(13) Thus, the amino acid assignments of the 61 coding triplets and the 3 termination triplets have remained the same if we disregard the few exceptions in maternally transmitted chloroplast and mitochondrial DNA.

The importance of brain size for digital processing becomes apparent from a comparison with computers. Information encoded in a binary code like Morse telegraphy requires very long sequences compared to the same text encoded in the English alphabet. The 4-letter word "ship", for example, requires 13 dots and dashes plus 4 spaces in Morse. An even more dramatic compression occurs, as we move from the 4-symbol code of nucleic acids to the 20-symbol language of proteins. In order to appreciate the significance of this compression, a few observations about translation might be helpful in this connection. The 64 coding triplets employed in translation are an interesting illustration of the Shannon redundancy. For, a non-uniform genetic code of 20 triplets or 16 duplets and 4 triplets would be much less error-resistant, quite apart from the fact that code words of unequal size (as in Morse) would be much more difficult to handle for the ribosomal translation machinery. Another consequence of the redundancy of the genetic code is the unidirectional flow of information from nucleic acid to protein immortalized in Crick's Central Dogma.(14) As pointed out by Yockey,(15,16) Crick's intuitively conceived dogma is a fundamental theorem and the consequence of Shannon's mathematical coding theorems. It applies to any communications system in which the entropy (or number of letters) of the sending alphabet (64 triplets) exceeds that of the receiving alphabet (20 amino acids). For the same reason, knowledge of a protein's amino acid sequence doesn't allow us to deduce its DNA coding sequence.

An even more fundamental property of the genetic code appears to have been overlooked: the compression of information and the digital to analog conversion produced by conferring specific three-dimensional patterns to proteins. As already mentioned, codes with few letters, the binary code of computers at the extreme, produce very long messages compared to codes with a larger number of letters. Imagine the length of a book written in binary code with one in the English alphabet. Apart from its volume and weight, the book written in binary code, would also be hard to read because such a code is unsuitable for pattern recognition, as will be discussed later. Digital computers, in contrast to the human brain, do not mind long sequences because their nanosecond processing speed is extremely fast compared to the millisecond duration of synaptic transmission. To make up for this handicap, the brain relies on massive parallel processing made possible by extreme miniaturization not achievable in silicon technology. In biology, the limit to this miniaturization is ultimately dictated at the molecular level by the size of the proteins. The enormous compression accomplished by the expression of DNA information in proteins becomes apparent, not so much from the relatively moderate linear compression, but most of all from a comparison of the volume of a protein with that of the coding region of its gene. The resulting reduction in volume is in the order of 100:1, but becomes even larger if we include the regulatory sequences of the gene. Although it seems to have escaped notice, it is obvious that, without this DNA-to-protein compression, the cell volumes and organ dimensions, including the size of the brain, would have to expand to unrealizable proportions. Hence, the invention of genetically encoded proteins for the construction, operation and evolution of multicellular organisms was a biological necessity. The genetic code was the magic formula.

The bulkiness of the chromatin fiber relative to its protein products is in turn dictated by its two different tasks: to serve as a stable information storage library and to replicate itself. The digital 4-letter code of DNA achieves both of these objectives optimally.(17) It follows from these considerations of molecular biology not only that brain size is essential in determining the power of digital processing but also why, compared to the most advanced silicon chip computers, the brain can pack so much more information and computing power into the modest 1,500 cc of our skull. Kurzweil has calculated that the computing power of a $1,000 personal computer, expressed as the number of calculations per second, is equivalent to that of a fly's brain.(18) He predicts that, at the present rate of accelerating progress in hardware design, the computing power of a $1,000 PC of the neural net type will in 2020 match that of the human brain.

Portmann: importance of early birth for neural development

Of particular interest here is the fact that the neural development for language occurs postnatally. The general importance of this was first recognized by Adolf Portmann in his pioneering studies more than half a century ago.(19) Emphasizing the significance of the small brain size at birth, Portmann argued that the sensory input provided by the extra uterine environment was essential at this stage of brain development. With regard to the acquisition of language as well as of other brain functions, he also pointed out, what has since been amply confirmed, that without this sensory input during a limited time period, the cognate functional neural networks cannot be formed. While this requirement for a postnatal phase of sensory input is shared by the cerebrally more developed mammals and birds, which Portmann called secondary "Nesthockers" (altricial), the duration of this development is much longer in humans.

Evidence from comparative anatomy and paleoanthropology

The time of origin and nature of pre-digital hominid communication has been controversial and the subject of intense speculation, especially in relation to brain size. The evidence available from paleontology, comparative anatomy, anthropology, linguistics and embryology may be assembled into the following picture. The australopithecines were the oldest, still apelike ancestors, who walked upright, as documented by the Laetoli footprints, sometime between three and four million years ago. Best known is the 3.2 million-year-old partial skeleton of Lucy,(20) which probably was that of a male.(21) The preserved knee-joint confirms the upright posture. The brain size of the individuals of this group was a little over 400 cc and grew to over 530 cc (*A. robustus*) during the next two million years. They overlapped in time for a million years with the first representatives of the genus *Homo,* classified as *H. habilis*, who lived from 2.4 to 1.6 million years ago and had a mean brain size of 650 cc and an upper limit of close to 800 cc (*H. rudolfensis*).

Compared to the great apes, the brain size of the early australopithecines does not appear to be significantly larger and only slightly increased in the late *A. robustus* (530 cc vs. 390 cc in an adult chimpanzee). If corrected for body weight, however, the brain volume of the australopithecines surpassed that of the great apes by 50%.(22) Hence it is not clear whether bipedalism required a larger brain or may have promoted it. With the appearance of *Homo* about two million years ago, the brain size in *H. habilis* grew another 50% to 650 cc, followed by another addition of about 400 cc in *H. erectus* during the next million years. One of its earliest specimens, *H. ergaster* WT 15000, the spectacular, 1.6 million year old nearly complete skeleton of the Turkana boy discovered by Richard Leakey(23) had a brain volume of 880 cc at the age of death at about 9(24) and would have grown to over 900 cc at adulthood. The skeleton is that of a slender youth, very similar to that of a modern boy of about 15. As an adult, he would have stood at least 6 feet. At the time of his death and for some 100,000 years afterwards, three different hominid developmental stages, *A. robustus, H. habilis* and *H. erectus*, were all living simultaneously in Africa, not so remarkable in the light of the low population density. During the following million years, until about 600,000 years ago, the endocranial capacity added only about 15% to 1043 cc in *H. erectus pekinensis*. The final value 1200–1500 cc was reached about 500,000 years ago with the emergence of modern humans in the form of *H. heidelbergensis* and later forms of Neanderthals, as well as archaic *H. sapiens* in Africa. Martin made a well-documented case for the view that brain size grew steadily over the past 3 million years driven by the gain in general intelligence made possible by the adaptation to a more energy-rich diet.(22)

In view of the impressive ability of apes like Kanzi to respond to human sign language and to understand simple spoken communication, we cannot stop thinking how much greater these communicative abilities would have been in hominids with twice their brain size. Yet, there is also a large gap between the ability to learn by human directive and the spontaneous development of an independent intra-species communication system based on phonetic articulation, especially in the absence of the required vocal equipment.

The human vocal tract anatomy differs from that of apes and the australopithecines by the much lower position of the vocal box in the neck, which prevents simultaneous breathing and swallowing. In addition, the shape of the bottom of the skull, the basicranium, flat in apes and australopithecines, becomes arched in humans, a process that began with *H. erectus* 1.6 million years ago and was fully developed in archaic *H. sapiens* 300,000 to 400,000 years ago.(25) This evolutionary history is again reflected in human postnatal development. At birth the location of the larynx high in the neck permits breathing while nursing. Only at one and a half years, coincident with early speech, does the vocal box begin to migrate down the neck to reach its final destination at age 14, a process reflected in the changing sound quality of the voice.

Most decisive for both the increase of brain size and the acquisition of language in humans is the early birth at a premature stage of CNS development. The bottleneck in prenatal brain expansion is the size of the female birth canal or pelvic opening, which accommodates the head of a human infant with a cranial capacity of 400 cc and appears not to have changed during hominid evolution. As already pointed out by Portmann,(19) this corresponds to the skull size of an adult chimpanzee. But in contrast to the ape, which is born with a brain volume of less than 200 cc, the human child, born with a brain volume of 370 cc, reaches the nearly fourfold larger adult size of 1400 cc at the age of 7. Even greater is the functional difference: the nervous system of the chimp baby is fully developed with regard to motility and muscle control and interaction with its mother, while the human infant is totally helpless, and its nervous system cannot reach functional maturity without exposure to the environment, which includes interaction with mother and other members of the family. The most important tool of this interaction is language.

The extrauterine brain size doubles during the first year, when bipedalism and most of the analog mental faculties are learned, and reaches 90% of its final volume at the end of four years, which coincides with the onset of the explicit digital mastering of language in the form of reading and writing. One of the most striking aspects of the early analog phase of development in children is the creation of an internal fantasy world with no strict relationship to the external reality, as expressed in a sheer insatiable appetite for fairy tales and fancy stories. In fact, stories hold such a fascination that they appear to the listeners to be more of a reality than the external world. Moreover, they insist on hearing the same stories over and over again and object when they are not precisely recounted. Given the fact that stories are pure language, this urge of the developing brain must be an important exercise in syntax, memory and social context, in short: the world of thought, reflection and introspection created by language. The same manifestation is seen in the tireless efforts and infinite patience aimed at muscular control of hands and fingers in coordination with vision during play with toys, especially those requiring the assembly of interlocking parts. It is known from neurobiology that such repetitive exercises are the method to program neural networks, an activity that artificial intelligence has been able to reproduce with computer networks with modest but growing success. The contrast to young apes, which are born with the dexterity of adults is most striking and emphasizes dramatically the significance of the early birth and extrauterine development of the brain in humans.

How and when this development got started has been a matter of controversy. Since the dimensions of the birth canal have not changed during the past 5 million years, the brain volume of about 400 cc, corresponding to that of the human newborn, is the limiting factor for intrauterine brain growth. In chimps, the brain is fully functional at birth with respect to most body functions, and the postnatal 2- to 3-fold increase in brain size reflects in part the overall body growth while at the same time allowing neural circuits to develop in response to the environment, as reflected in the learning phase of play-acting, characteristic of young mammals. By contrast, the human infant's head has already reached at birth the maximum size permitted by the birth canal and must be born in a premature state to allow a fourfold postnatal increase. From the analogy to the situation in chimps, and from the assumption that in the early hominids the brain volume at birth was about half that of the adult and fully functional, it follows that, up to a brain capacity of 400 cc at birth and 800 cc fully grown, the size of the birth canal did not necessitate the birth of babies with developmentally immature brains. Hence, *H. habilis* and early *erectus* with an adult brain capacity of about 800 cc could have reproduced without the necessity for extrauterine brain development. This would place the change toward raising helpless, cerebrally immature infants at the border between *H. erectus* and archaic *H. sapiens* some one million years ago.

The enormity of this revolution is nothing less than staggering. In view of the high price in terms of total dependency and need for protection, the selective advantages of being born in what amounts to a fetal state are not obvious. The success of the experiment implies that the climatic and social conditions were favorable and early humans had progressed to a state of being secure from predators and becoming hunters themselves, able to satisfy the more demanding nutritional requirements(22) of the growing brain.

A greater dependency on less-abundant food sources would in turn have a limiting effect on population size. For throughout history food has been the only factor limiting the growth of populations. Before the advent of digital language some 100,000 years ago, the population density of the genus *Homo* was minuscule and hence not conducive to the spread of communication tools. Yet as the postnatal brain grew and with it memory expanded, some sort of proto-language may have evolved, probably similar to the early analog phase of children's speech. Why according to all accounts digital processing took over only between 100,000 and 50,000 years ago, or some 200,000 to 400,000 years after the brain had reached its present size is still a mystery. One possibility is that during this time important internal changes not reflected in external brain size occurred. As a result, further brain growth stopped abruptly, because now the selective advantages of the newly created oral tradition could not possibly be matched by further brain growth whose cost would have become prohibitive. It is attractive to speculate, therefore, that, as brain growth in the Neanderthals evidently failed to stop, their larger brains had not acquired digital language, and this was the reason for their extinction.

Most experts think that the digital processing characteristic of all modern languages evolved within a relatively short period. The inference that it resulted from a rather sudden genetic reorganization of the brain(26) does not exclude the relatively slow accumulation of mutations that in a last step brought about the manifestation, which appeared so abruptly. Its late arrival in evolution is consistent with the observation that in ontogeny it coincides with the stage of maturation at about age four, or two years after language acquisition, when the brain has reached 90% of its final size. It is interesting in this connection that the memory of adult people rarely extends to events before the age of four, quite in contrast to children of four and younger who precisely recall experiences reaching much further back. It is tempting to explain this early childhood amnesia as the result of a digital reorganization that wipes out earlier analog memories.

Structural similarity of proteins and human language

The structural similarity of language and proteins as mediators of both digital and analog information is striking and deserves further exploration. In both cases their technical transmission obeys Shannon's theorems. For genes, the subdivision into exons, introns, into motifs determining all possible interactions between proteins and nucleic acids, are impressive illustrations of the economy of the digital principle. In language, recombination of syllables into words and of words into sentences has its counterpart in the variety of patterns achieved by different combinations of motifs and domains at the nucleic acid and protein level, and in the formation of the larger units of protein subunits, homo- and hetero-oligomeric proteins, and finally in the pathways or sentences formed by the association of these elements into a functional ensemble with a specific objective within a still larger structure. The same hierarchical principle is used in the construction of the phenotype by cells forming tissues that in turn are assembled into the organs making up the various systems of the organism. While human speech, like acoustic communication in the animal kingdom, is still an analog function based on pattern recognition, its new richness and cognitive power was made possible only through the introduction of the digital principle embedded in the syntax of all human languages. How and when did it arise?

Evolution from analog to digital

The primacy of the analog in brain function, because it reflects the oldest evolutionary history, is evident in most human activities. It is not surprising therefore that the same trend is seen in science; outstanding examples are pre-Mendelian genetics and Pauling's analog antibody model preceding the digital one of Jerne. The same principle is observed in the acquisition of language by children. I imagine that phonetic communication began with simple sound patterns like mama and papa. These pre-human sound patterns were associated with objects and activities, much like specific birdcalls warn from impending danger. In this type of protolanguage a limited vocabulary may have been preserved over generations by oral replication until the rather sudden discovery by chance and experimentation of the variety of sound patterns that can be produced combinatorially and exploited by assignment to things and activities. The question of the triggering event will probably never be answered; it could be one of the phase jumps characteristic of adaptive evolving systems.(27) Such an accident was bound to happen after the brain had attained the required structural complexity.

Chomsky concluded from his studies that the similarity of the syntax of all known languages must have a genetic basis, a concept that revolutionized linguistics.(28) The complexity of the underlying neural networks would of necessity require a very long time to evolve. It is not surprising, therefore, that recently discovered stone-age societies who lack written communication, nevertheless are fully conversant in their own digital language and capable of learning digital script.(29)

From analog to digital script

A paradigm shift occurred with the invention of writing by the Sumerians some 5,500 years ago, the most important event in human history. It was repeated independently in Egypt about 200 years later. In both cases, the original script was analog or pictorial. In the case of the cuneiform analog script, the extensive use of writing by the Accadian merchants in Babylon promoted 2000 BC the rapid transformation to a digital form which reached its final phonetic form in the cuneiform representation of syllables.(30) In Egypt, where writing was the privilege of priests, the evolution to digital writing took longer and remained a curious mixture of pictorial and digital signs, which misled Jean-François Champollion for a long time to believe it was analog. Zoega 1797 suggested that the hieroglyphs were letters,(31) a hunch confirmed 1802 by the Swedish diplomat Johan David Akerblad who identified 14 signs.(32) The decisive digital breakthrough came 1814, when the English physician and physicist Thomas Young, who knew 13 languages and used a scientific approach similar to modern-type cryptological methods, succeeded in deciphering the first cartouche.(33) Only then was Champollion 1821 able to finish the job, but never gave credit to Young for putting him on the right track.(34) The final digitalization of writing was accomplished much later around 800 BC in the Greek alphabet with the resolution of the syllables into their ultimate phonetic components.(30)

Acquisition of language by children

The strongest evidence for the idea that the digital structure of human language is a late development in evolution, and the key to the origin of modern *H. sapiens*, comes from comparative studies of language acquisition in apes and children.(35) Experiments with young gorillas and chimpanzees, especially the talented bonobos, have revealed that these apes are capable of an extremely limited understanding of human language and of analog communication by sign language. Claims that they can even be taught to read analog symbols for use in a limited vocabulary and to combine the symbols in a primitive syntax have been seriously questioned.(36) The apparent understanding of some of the American Sign Language (ASL) signs was, according to Goodall, based on gestures that she had observed the chimps using in the wild rather than an ability to learn and comprehend the essentially digital structure of ASL. So it is not surprising that they are totally incapable of learning digital script. Comparing these efforts to teach human communication to apes and children, it is striking that children are eager to learn and derive great enjoyment from it, while apes must be coaxed and bribed with rewards. While this lack of enthusiasm appears not to apply to tasks in their natural habitat, the significant difference is that humans are not subject to similar constraints; in fact it is precisely this lack of specialization that has spurred brain growth throughout evolution.

In children, language acquisition occurs in several stages, only the first of which has similarities to certain observations with apes. The recognition of the sound patterns of speech and its comprehension begins soon after birth and long before speech production, which begins only after the first year. The accumulation of a productive vocabulary follows and results in the ability to produce syntactically correct sentences by the end of the second year. This development is accompanied by an acute ability of visual pattern recognition beginning right after birth. For example, my granddaughter Isabel was able at the age of 18 months to recognize and name in a puzzle of the states of the US every piece not only from its position on the map but also as single pieces from their shape. At early age, recognition of shapes precedes that of pictures in solving the puzzle. Still, children are incapable of learning to read digital script before the age of four. Clearly, the required digital processing power of the brain is a very late event both in human evolution and postnatal development. If this interpretation is correct, we would expect that children could be taught analog writing at a much earlier age. Indeed, some Chinese children begin to learn their analog script at the age of about 18 months.

In summary, acquisition of language in children may be divided into three phases, (i) an early analog phase of learning to recognize the sound patterns of speech and their meaning, (ii) an intermediate phase of producing speech, and (iii) a late phase of learning to write and read digital scripts. Phase (i) is largely analog and, in contrast to phase (ii) and (iii), shared to some extent with other mammals and birds. All three phases appear to correspond to stages in evolution, which in turn reflect successive stages of postnatal brain development, probably representing the maturation of anatomically distinct neurological structures. Phase (ii), which forms the bridge between the purely analog and digital, may therefore already require some digital processing not available to animals. It may also correspond to the development stage of the early forms of *Homo*.

The intimate connection between phase (ii) and (iii) is also apparent in the process by which children learn how to read. One of the methods that was popular a generation ago, taught reading by the analog method of recognizing word patterns. The limitation of this method is that children were unable to cope when confronted with a new word, an obstacle not encountered with the much slower, more general and demanding method of digital synthesis or phonetic spelling. Because of the much greater speed of pattern recognition and the associated memory function, it is not surprising that this process takes over each time a new word has been digitally decoded.

Somatic and genetic neuropathology of speech

Additional evidence in favor of the interpretation of the observations presented here comes from neuropathology. It is well known that victims of strokes or of accidental or hereditary brain injuries often have lost the faculty of speech. Of particular interest in this connection is the case of the famous Italian filmmaker Antonioni who after a stroke was unable to speak, read and write while perfectly capable of understanding all speech. To communicate, he had to resort to drawing pictures. Thus, the damage inflicted on his brain produced a block in the production of the digital elements of language while leaving the structures for analog communication intact. This example illustrates my point well because Antonioni's inability to speak was not the result of a defect in the motor function as in Broca's aphasia. It is interesting in this connection that in aging the late-acquired digital functions are the first to get lost, as evident from the difficulties remembering names while still recognizing the cognate faces.

Certain lesions, like those causing Wernicke's aphasia, may affect only part of the digital processing network with the result that these patients speak nonsense, invent new words and scramble the order of words in a sentence, as if the digital machine was still working but not producing any meaningful patterns. Inasmuch as digital functions necessary for speech and, at the highest level, for reading, most certainly involve multigene networks, we would expect that mutations in these genes would cause partial deficiencies such as the loss of the ability to read without affecting speech. Similarly, in severe cases of dyslexia, those afflicted are unable to read and write nor can they be taught. Yet, despite this digital deficiency, they converse normally as long as the subject matter discussed does not reach a level of even modest abstraction requiring digital processing. If these digital defects are acquired rather than genetic their causes remain elusive. An interesting exception is the rare Williams syndrome, which has been traced to mutations in genes involved in calcium metabolism.(37) Children suffering from this syndrome learn to speak normally, but read and write only poorly and have limited comprehension.

Another genetic impairment of grammar was found to be transmitted like a dominant mutation in a single gene.(38) Those afflicted are unaware that there are general rules of English grammar for producing plurals and tenses and must learn each word as a separate lexical item, e.g., that the word "books" refers to several books, etc. What makes this mutation so significant, is that, rather than affecting digital processing per se, it interferes with parts of the circuitry responsible for generating particular generalizations. The identification of the gene product, most likely a transcription factor, is of paramount interest, as it is expected to give us access to the genetic network controlling grammar.

Selective advantage of digital acoustic communication

The idea that the digital structure of all languages is the distinctive feature of humans, that its acquisition is in fact the crucial event in the genesis of *H. sapiens sapiens*, is supported by a number of facts and observations not previously viewed in this context. Given the selective advantages of acoustic communication, the digital mode is most likely dictated by neurological economics. While the auditory discriminating power is enormous, as reflected in voice recognition, the repertory of sounds available to the organs producing speech seems to be much more limited. Even more serious is the problem of learning to produce 10,000 different sounds corresponding to an equivalent vocabulary of different meanings as in the case of the Chinese characters. The solution, of course, is to reduce the number of patterns to be produced and recognized by combining, according to the digital principle, a much smaller number of sounds into different sound patterns consisting of the hierarchy of phonemes, syllables, words and sentences. A necessary condition is, evidently, the production of sound patterns consisting of discontinuous, distinguishable sound elements.(39)

The problem and its solution is best illustrated by considering that the number of letters in a phonetic alphabet is between 25 and 50, which corresponds to all the different sounds that are easily produced by our verbal and recognized by our auditory sensory equipment. Without making use of the combinatorial digital principle, this would limit our vocabulary to between 25 and 50, or at most, a few hundred words. By combining one vowel with two consonants as in mama and papa, we can generate 3^4 or 81 different 4-letter words, or 16 useful ones if we eliminate pm and mp as too difficult to pronounce and restrict ourselves to combinations of ma, pa, am, and ap as in pama, amam, apap etc. In English only two of the 81 theoretical or 16 of the phonetically realizeable 4-letter words, namely mama and papa, are used if we disregard maam, which by introduction of the guttural stop becomes ma'am for madam. The enormous number of sound patterns made possible by the digital principle is obvious from this example. A further extension of the digital principle is to combine syllables or phonemes into words and words into sentences. Several conclusions follow immediately. (i) The digital principle is used to generate a large number of sound patterns, which are perceived by the analog mechanism of pattern recognition, a process which in turn is facilitated by the limited number of discrete sound elements employed. (ii) Because of the large number of possible patterns, those selected in a language are an extremely small proportion of the total. The fact that certain words have been conserved within a large spectrum of languages as in the case of mama and papa, would therefore point to a common origin of these languages.(40,41) (iii) Since this selection process is the result of evolution and history, many different languages are expected to evolve in analogy to the process of speciation in multicellular organisms. (iv) The digital structure of language is a condition for its replication by oral tradition or, in written form, for its copying, storage and processing in computers and, presumably, in the brain. I do not want to imply, however, a structural analogy between computers and the brain. Rather, I wish to emphasize that both use the digital principle for similar objectives.

As our understanding of biology progresses, we discover a growing number of examples in which biology has anticipated human technology. A striking illustration in information processing is the analogy between a tape-operated linotype printing press (now replaced by fully electronic machines) and the translation of the messenger tape on the ribosome in protein synthesis.(42) We are now rapidly entering the stage where human technology deliberately attempts to copy biological mechanisms. The first successful applications of this "reverse engineering"(18b,43) have recently been accomplished in neurobiology. It is therefore not surprising that structure and evolution of language are a reflection of information processing at the molecular, cellular and multicellular level.

Transformation of digital information into analog function

Although not mentioned in textbooks, it is hard to imagine how life could have arisen without the invention of the digital principle in the form of RNA and DNA. The translation of the

4-letter DNA- and RNA-code into the 20-letter amino acid code follows strictly Shannon's theorems. The evolution of the protein complexity parallels that of human language in that it involves the digital-to-analog translation from nucleic acid into the enormous variety of three-dimensional structures of proteins. In this process, the digital information in DNA is transformed into the analog information of protein function. Again only a tiny fraction of the possible DNA and amino acid sequences are selected to produce useful proteins. And because neither the folding process nor the protein language are understood, the analog information remains a mystery even when the DNA sequence is known. We are in the same situation as a reader confronted with a text in a foreign language. To learn the analog language of proteins, we must correlate DNA sequence motifs with protein-folding and -interaction patterns in a process analogous to that of a child learning a language by associating analog sound patterns with objects and activities. In biology this learning process has given rise to molecular genetics and the new science of genomics, which attempts to associate a DNA sequence with a particular function, e.g., by tracing known changes in DNA sequence to observable changes in function. Here the situation is further complicated by the fact that not all DNA sequences encode proteins, yet still are involved in protein function.

It is interesting in this connection that John von Neumann in his 1948 lecture on "The General and Logical Theory of Automata" made some very astute observations on biological systems. Comparing analog and digital computers, he predicted the importance in living systems of feedback loops of alternating digital and analog reaction chains. He was convinced that the complexity of brain function required digital processing.(44)

Mendel's discovery of digital genetics

In retrospect and our modern terminology, Mendel's 1866 demonstration that genetic encoding is digital,(45) substantiated a century later at the molecular level by the DNA structure of Watson and Crick(46) and by its translation into individual, unique amino acid sequences, will forever remain one of the greatest discoveries in biology. Mendel's failure to make an impact during his lifetime and for half a century after his landmark paper(47,48) is another illustration of the fact that comprehension requires compatibility between the patterns of the sender and the receiver. The greater the revolutionary character of a discovery, the less it will fit the prevailing landscape. That his paper was published in a relatively obscure journal, while not exactly helpful, would not have prevented its immediate success, as was accorded the DNA double helix, had the time been similarly ripe. But none of Mendel's contemporaries was thinking in terms of a digital principle of information storage and transfer. It took the genius of this monk, who was trained as a physicist, to apply this type of thinking to biology, most unusual for that time. Particularly revealing is his own thought about the significance of his work expressed in a letter to Naegeli (the arrogant, uncomprehending, overbearing and consistently wrong pope of the field): "*I knew that the results I obtained were not easily compatible with our contemporary scientific knowledge, and that under the circumstances publication of one such isolated experiment was doubly dangerous, dangerous for the experimenter and for the cause he represented*".

Genetic mechanisms of postnatal phenotype modification

The radically new situation that so impressed Portmann more than half a century ago, is the inheritable ability of the phenotype to be modified by the environment. Or, the ability of the environment to actively direct structural features of developing tissues toward a certain purpose. Lamarck sneaking in through the backdoor?

The only other system with the capacity of learning in a developmental dialog with the environment is the immune system. To explain the specificity of antibodies, the accepted model in the forties, Pauling's instructive theory was purely analog by imagining that the antigen–antibody complementarity resulted from the protein chain of the immunoglobulin folding around the antigen. The first to challenge this instructive theory with a digital model was Jerne. As I was sharing his office at the State Serum Institute in Copenhagen from 1948 to 1950, I became aware that he was thinking in terms of information processing by a mathematical approach. His first clue came from the observation that both the concentration of antibodies and their affinity for the antigen increased with time after immunization.

This led him to postulate the existence of preformed antibody molecules of a given specificity or fit, which was improved by a postulated somatic mechanism of repetitive selection,(49) now known as affinity hypermutation. Extending this idea, Burnett proposed that the selection operated at the cellular level, an idea now known as the clonal selection theory.(50) The final experimental proof was again given by Jerne with the ingenious hemolytic plaque method which he had conceived on paper at his WHO desk job in Geneva and realized experimentally at the University of Pittsburgh in 1961.(51) Jerne, in his classical paper "Towards a network theory of the immune system" was also the first to point out the close analogy between the immune and the nervous system: *"(the) immune system, when viewed as a functional network dominated by a mainly suppressive Eigen-behavior, but open to stimuli from the outside, bears a striking resemblance to the nervous system . . . Like for the nervous system, the modulation of the network by foreign signals represents its adaptation to the outside world. Early imprints leave the deepest traces. Both systems thereby learn from experience and build up a memory that is sustained by* *reinforcement and that is deposited in persistent network modifications, which cannot be transmitted to our offspring. These striking phenotypic analogies between the immune system and the nervous system may result from similarities in the sets of genes that govern their expression and regulation*''.(52)

Later, in his equally elegant Nobel lecture entitled ''The generative grammar of the immune system'' Jerne discusses some of the analogies between the immune system and language.(53) Referring to Chomsky's definition of generative grammar(28) as a device containing a central syntactic component, a phonological component and a semantic component, Jerne tries to apply this concept to protein structures. Thus, he equates the rules of syntax to the rules determining the folding pattern of the polypeptide chains, but finds it harder ''*to find an analogy to semantics: does the immune system distinguish between meaningful and meaningless antigens*?'' Leaving this question open, he turns to the inheritable structure of the immune system and points to the generative capacities of the proliferating lymphocyte as they turn on the somatic mutations in the DNA segments encoding the variable regions of the antibody polypeptides. He concludes with ''*the miracle that young children learn the language of any environment into which they are born. The generative approach to grammar, pioneered by Chomsky, argues that this is only explicable if certain deep, universal features of this competence are innate characteristics of the human brain. Biologically speaking, this hypothesis of an inheritable capability to learn any language means that it must be encoded in the DNA of our chromosomes. Should this hypothesis one day be verified, then linguistics would become a branch of biology, and the humanities, perhaps, some day part of the sciences*.''

Jerne is right in comparing those structural similarities of the immune system and language that I have defined as digital, and he realizes that the unsolved problem is how to get from the genotype to the phenotype, across the chasm separating unknown neural structures from the actual output expressed as speech. The trouble starts, however, with his attempt to equate the syntactic component with the rules of polypeptide chain folding. As explained earlier, we are confronted with a digital-to-analog conversion, which joins the syntactic to the semantic, as it implements syntactical rules to express semantic meaning. Similarly, in protein synthesis the analog information or semantic meaning created by chain folding is the resulting three-dimensional shape, which functions in pattern recognition. All biological information transfer is based on pattern recognition, as it involves the physical interaction of complementary three-dimensional surface structures of the protein in question with other proteins, as in antigen-antibody binding, or interactions resulting in the catalytic modification of metabolites by enzymes, interactions of transcription factors with DNA in gene activation or interactions with receptor proteins in signal transduction. Exactly how all of these interactions are integrated to produce the phenotype with the stunning precision illustrated by the similarity of twins sharing copies of the same DNA is still a deep mystery, which will probably forever elude us. No wonder that Chomsky wants to stay away from this biological trap.

To return to our original question, we may now summarize what the immune system teaches us as a system capable of learning from the environment. The lesson is that, in seeming violation of the central dogma, an invading foreign molecule acts as the signal that provokes the synthesis of proteins capable of specifically combining with the invader. Elucidation of the molecular biology of this response has shown that the non-Shannon information or pattern of the invading antigen is not transferred to the DNA by a reversal of protein synthesis, but that the foreign pattern triggers the transcription of pre-existing DNA segments. To cope with the incredible diversity of about 10^8 possible molecular shapes, it would not have been feasible to store permanently in the genotype that many antibody genes, especially in view of the 100:1 compression discussed earlier. What nature has invented instead, in a trick that bypasses the genotype–phenotype barrier, is an information-generating machine that, in the maturing lymphocyte, assembles on command by genetic recombination of several separate genes a combinatorial multiplicity of new genes, which encode the diversity of some 10^{15} antibody molecules. Since one antibody-producing cell or clone, called B-lymphocyte, can synthesize only one type of antibody, the body needs at least 10^8 different B-lymphocytes, each representing a combining pattern, which it displays on its surface in the form of its specific antibody. Those antibodies on the surface of the B-lymphocyte fitting the pattern of the foreign molecule, the antigen, trigger the clonal expansion of the cell producing it, which now begins the mass production of the desired immunoglobulin at the rate of 2000 molecules per second. During this process, another information-generating machine in a fine-tuning operation grinds out point mutations corresponding to three positions of the antigen-combining site, the hypervariable sites, a process called affinity maturation because it allows another round of selection for the best fit. After elimination of the invasion, antibody synthesis returns to basal levels as a result of programmed cell death. The demobilization, however, is only partial because the information of the survived attack is stored in memory cells that serve as a kind of RAM and can be rapidly activated by a renewed encounter with the same antigen. The ensuing secondary response is both more vigorous and specific, a fact exploited in immunization.

An interesting illustration is the problem presented by viruses like HIV that attack specifically the immune system. What makes this virus so deadly is its strategy to pitch its own information-generating machine against that of the host. The virus escapes destruction by mutating frequently to new patterns not recognized by memory cells. This game of attack and counter-attack often continues for years until the immune system of the host is exhausted.

The important lesson here is that nature has developed, at the level of the phenotype, mechanisms of learning in response to the environment based on directed analog modification of DNA, a process aptly termed ‘‘adaptive Lamarckism’’ by Maynard Smith.(54) The same principle applies to the postnatal development of the brain although much less is known about its molecular biology than in the case of the immune system. At present nearly all the information must be inferred from observing the development of language acquisition in children, rare mutations affecting language, and from attempts to mimic learning with computer nets. One of the central questions has been the Chomsky conundrum that children learn to speak correctly without knowing the complex structure and rules of grammar because they inherit this knowledge in the form of neural circuitry, also called universal grammar (UG) that generates the phenotype of actual speech. This view regards the inherited generative grammar as a Shannon-type of digital information devoid of any semantic or analog meaning.(55) From a formal point of view, this concept is unassailable, especially since Chomsky makes no attempt to explain the mechanism leading from genotype to phenotype.

Possible approach to a solution of Chomsky's conundrum: postnatal formation of logical neural networks

The concept that I wish to propose here is that the syntax studied by linguists is an abstraction that is not definable by specific static brain structures. Rather I believe what is unalterably encoded in the DNA are the instructions for making a neural network capable of being dynamically programmed by sensory input during a postnatal period of brain growth and development. Networks have an internal logic that sorts and imparts structure on the sensory inputs by testing them through experience. Exposure to language develops comprehension by repetitive inputs in association with specific experiences. The rules are not preformed as Chomsky appears to imply, but arise as a network function in trial-and-error exercises. Any network of this complexity will deduce what appears as rules from being exposed to many examples. This is reflected in the fact that it is easier to grasp abstract concepts from examples than from analytical deductions. Even in mathematics, the most abstract science, great ideas have been conceived by intuition, only to be proved in a time-consuming, laborious process, sometimes requiring centuries, as in the case of the proof for Fermat's last theorem. While the molecular mechanism of network programming in the brain is still unknown, I expect that, as in the case of the maturing lymphocyte, some form of somatic genetic or post-transcriptional recombination produces a great diversification of the proteins involved in axon guidance(56) and synapse formation, and that the auditory signals of spoken language in combination with visual clues are the selective agents in the formation of specific circuitry. This idea is consistent with the observation that some people, who lost speech as a result of brain damage, are able to learn it again because their genetic arsenal for reconstructing the damaged circuits is intact. Their situation is different from people with mutations of the kind described by Gopnik(38) who are genetically impaired in their learning ability.

It is a general feature of evolving self-organizing complex systems, consisting of a network of multiple linked feedback loops, to form a dynamic equilibrium with the environment. The stability and ‘‘intelligence’’ of such systems is greatly enhanced, as the numbers of its components is increased. Recent observations of such diverse examples as the internet, free market economics, and insect societies suggest that such systems develop their own logic and respond similarly to external inputs.

In conclusion, I believe that the ideas and supporting evidence presented in this paper strongly favor the view that digital processing as a result of postnatal brain expansion is the revolutionary event responsible for our digital language. The digital structure was the simplest solution for acoustic communication allowed by the available anatomic structures for producing and perceiving auditory signals. Acquisition of a digital language with its open-ended possibilities is a development of the last 100,000 years that defines modern *H. sapiens* and sets him apart from all earlier forms of life.

Acknowledgments

This paper is in part based on a series of lectures I gave as visiting professor in the Departments of Cell Biology and Environmental Sciences at the Federal Institute of Technology in Zurich, Switzerland during the winter 1992/1993. Special thanks go to Theo Koller for generous hospitality and stimulating discussions during that time. I thank Markus Noll, Rebekka Cann, Felix Boehm, Tom Humphreys, Norton Zinder, Günter Blobel for a critical reading of the manuscript, discussions and suggestions. I am greatly indebted to Frederick Newmeyer for criticism and valuable advice.

I dedicate this paper to the memory of my friends and mentors Arthur Stoll (1887–1971), Friedrich Dürrenmatt (1921–1990), Adolf Portmann (1897–1982), and Niels Kaj Jerne (1911–1994).

References

1. Shannon CE. A mathematical theory of communication. Bell System Technical Journal 1948;27:379–423 and 623–656.
2. Gell-Mann M. The quark and the jaguar. New York: W.H. Freeman & Co.; 1984.
3. Bresch C. Die Rolle der Sprache im Gesamtbild der Evolution. Nova Acta Leopoldina 1981;54:747–752.
4. Planck M. Scientific biography and other papers. London: Williams & Norgate Ltd. 1950. p 13–51.

5. Shannon CE. Prediction and entropy of printed English. Bell System Technical Journal 1951;30:50–64.
6. Shannon CE. An algebra for theoretical genetics. Ph.D. Thesis. Massachusetts Institute of Technology; 1940.
7. Thurber J. The unicorn in the garden. The Thurber Carnival. New Modern Library, Random House. 1931. p 268–269.
8. Tattersall I, Matternes JH. Once we were not alone. Scient Amer 2000;282:56–62.
9. Dunbar R. Grooming, gossip and the evolution of language. Cambridge: Cambridge University Press; 1998.
10. HE Gruber, JJ Vonèche. (Eds). The essential Piaget. New York: Basic Books Inc.; 1977. p 496.
11. Luquet H. Le dessin enfantin. Alcan, Paris; 1927.
12. Deacon DE. The symbolic species: The co-evolution of language and the brain. New York: W.W. Norton & Co.; 1997. p 373.
13. Eigen M, Lindemann BF, Tietze M, Winkler-Oswatitsch R, Dress A, von Haeseler A. How old is the genetic code? Statistical geometry of tRNA provides an answer. Science 1989;244:673–679.
14. Crick F. Central dogma of molecular biology. Nature 1970;227:561–563.
15. Yockey HP. Information theory and molecular biology. Cambridge: Cambridge University Press. 1992. p 95–108.
16. Yockey HP. Can the central dogma be derived from information theory? J theor Biol 1978;74:149–152.
17. Cullmann G, Labouygues JM. Le code génétique, code instantané absolument optimal. Compt rend Acad Sci III 1985;301:157–160.
18. Kurzweil R. The age of spiritual machines. New York: Viking Penguin 1999; (a) pp 102–106; (b) pp 118–128.
19. Portmann A. Biologische Fragmente zu einer Lehre vom Menschen. Basel: Benno Schwabe & Co. Verlag; 1944.
20. Johanson DC, Lovejoy CO, Kimbel WH, White TD, Ward SC, Bush ME, Latimer BM, Coppens Y. Morphology of the Pliocene partial hominid skeleton (A.L. 288-1) from the Hadar Formation, Ethiopia. Am J Phys Anthrop 1982;57:403–451.
21. Häusler MF, Schmid P. Comparison of the pelves of Sts 14 and AL 288-1:implications for birth and sexual dimorphism in australopithecines. J Hum Evol 1995;29:363–383.
22. Martin RD. Human brain evolution in an ecological context (52nd James Arthur Lecture on the Evolution of the Human Brain). American Museum of Natural History; 1983.
23. Brown F, Harris J, Leakey R, Walker A. Early *Homo erectus* skeleton from west Lake Turkana, Kenya. Nature 1985;316:788–792.
24. Smith BH. Dental development in *Australopithecus* and early *Homo*. Nature 1986;323:327–330.
25. Laitman J. The anatomy of human speech. Natural History 1984;92:20–27.
26. Bickerton D. Language and human behavior. University of Washington Press; 1995.
27. Eigen M. Phase jumps—the physical basis of life. Between Science and Technology. A Sarlmijin, P Kroes. (Eds). North-Holland: Elsevier Science Publ. B.V. 1990. p 119–153.
28. Chomsky N. Language and Mind. New York: Harcourt Brace Jovanovich; 1972.
29. Klein G. Proteus II-Carleton. In: Live now, Prometheus Books, Amherst, New York. 1997. p 85–164.
30. Lecours AR. The origins and evolution of writing. In: J-P Changeux, J Chavaillon. (Eds). Origins of the Human Brain. Oxford: Oxford University Press. 1995. p 213–235.
31. Zoega J. De origine et usu obeliscorum. Rome. 1797. p 465.
32. Akerblad JD. Lettre sur l'inscription égyptienne de Rosetta adressée au Citoyen Silvestre de Sassey, Paris, 1802.
33. Young T. Letter from W.E. Rouse Boughton to the Rev. S. Weston respecting some Egyptian antiquities, with four copper plates. Archeologica (London) 1814;18:59–72.
34. Andrews C. British Museum Book of the Rosetta Stone. London: British Museum Publications Ltd.; 1983.
35. Savage-Rumbaugh ES. Ape Language: From conditioned response to symbol. New York: Columbia University Press; 1986.
36. Pinker S. The language instinct. New York: William Morrow and Co.; 1994, p 338.
37. Bellugi U, Bihrle A, Jernigan T, Trauner D, Doherty S. Neuropsychological, neurological, and neuroanatomical profile of Williams syndrome. Am J Med Genet Suppl 1991;6:115–125.
38. Gopnik M. Feature-blind grammar and dysphasia. Nature 1990;344:715.
39. Zimmer DE. So kommt der Mensch zur Sprache. Heyne Sachbuch No. 19/310. München: W Heyne. 1986. p 18–19.
40. Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J. Reconstruction of human evolution:Bringing together genetic, archeological, and linguistic data. Proc Natl Acad Sci USA 1988;85:6002–6006.
41. Ruhlen M. The origin of language. New York: John Wiley and Sons; 1994.
42. Noll H. Polysome organization as a control element. In: Developmental and metabolic control mechanisms and neoplasia. 19th Annual Symposium on Fundamental Cancer Research. University of Texas M.D. Anderson Hospital and Tumor Institute, Houston, Texas. Baltimore: The Williams & Wilkins Co.; 1965.
43. Port O. Artificial vision. Bus Week Jan 31, 2000.
44. von Neumann J. The general and logical theory of automata. Hixon Sympos California Institute of Technology Sept 20, 1948.
45. Mendel G. Versuche über Pflanzen-Hybriden. Verh naturf Ver Brünn 1866;4:1–47.
46. Watson JD, Crick FHC. A structure for deoxyribose nucleic acid. Nature 1953;171:737–738.
47. Hartl DL, Orel V. What did Gregor Mendel think he discovered? Genetics 1992;131:245–253.
48. Sandler I, Sandler L. A conceptual ambiguity that contributed to the neglect of Mendel's paper. Hist Philos Life Sci 1985;7:30–70.
49. Jerne NK. The natural-selection theory of antibody formation. Proc Natl Acad Sci USA 1955;41:849–857.
50. Burnett FM. The clonal theory of acquired immunity. Cambridge: Cambridge University Press; 1959.
51. Jerne NK, Nordin AA. Plaque formation in agar by single antibody-producing cells. Science 1963;140:405.
52. Jerne NK. Towards a network theory of the immune system. Ann Immunology (Inst Pasteur) 1974;125C:373–389.
53. Jerne NK. The generative grammar of the immune system. EMBO J 1985;4:847–852.
54. Maynard Smith J. The idea of information in biology. Quart Rev Biol 1999;74:395–400.
55. Edelman GM. Bright Air, Brilliant Fire. New York: Basic Books Inc.; 1992, p 243.
56. Schmucker D, Clemens JC, Shu H, Worby CA, Xiao J, Muda M, Dixon JE, Zipursky SL. *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 2000;101: 671–684.